PCT/AU2004/000480

Australian Government

Patent Office
Canberra

I, JANENE PEISKER, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. 2003901717 for a patent by WESTERN SYDNEY AREA HEALTH SERVICE as filed on 10 April 2003.

COMMONWEALTH OF AUSTRALIA
AUSTRALIA
PATENT OFFICE

WITNESS my hand this
Eighth day of March 2005

JANENE PEISKER
TEAM LEADER EXAMINATION
SUPPORT AND SALES

AUSTRALIA

Patents Act 1990

Western Sydney Area Health Service

PROVISIONAL SPECIFICATION

Invention Title:

Identification of Streptococcus pneumoniae *serotypes*

The invention is described in the following statement:

IDENTIFICATION OF *STREPTOCOCCUS PNEUMONIAE* SEROTYPES

FIELD OF THE INVENTION

The present invention relates to molecular methods of typing *Streptococcus pneumoniae*, as well as polynucleotides useful in such methods.

BACKGROUND OF THE INVENTION

Streptococcus pneumoniae is a leading cause of morbidity and mortality causing invasive disease such as meningitis and pneumonia as well as more localised disease such as acute otitis media and sinusitis. Polysaccharide and protein-conjugate pneumococcal vaccines have the potential to prevent a significant proportion of cases. Effective protein-conjugate vaccines are particularly important because of the dramatic increase in prevalence and international dissemination of antibiotic resistant *S. pneumoniae* serotypes that commonly cause invasive disease in children (Hausdorff et el., 2001; Huebner, et al., 2000). However these vaccines protect against only the relatively small minority (Dunne et al., 2001; Hausdorff et el., 2001) of pneumococcal serotypes that most commonly cause disease. There is theoretical and limited empirical evidence that widespread use of these vaccines could lead to substitution of "vaccine" serotypes with other nonvaccine serotypes, against which the vaccines to not provide protection. Continued surveillance will be critical to monitor vaccine efficacy and changes in incidence and distribution of colonising and invasive serotypes (Hausdorff et el., 2001; Rubins et al., 1999). Any increase in disease caused by previously uncommon nonvaccine serotypes could necessitate a change in vaccine composition (Lipsitch, 2001).

S. pneumoniae comprises at least 90 serotypes, distinguished by capsular polysaccharide antigens. The capsular polysaccharide synthesis (*cps*) gene clusters for at least 16 pneumococcal serotypes have been sequenced and serotype-specific genes identified (Jiang et al., 2001; van Selm et al., 2002). The *cps* gene cluster contains genes responsible for synthesis of the serotype-specific polysaccharide including - except in serotype 3 - *wzy* (polysaccharide polymerase gene) and *wzx* (polysccharide flippase gene). At the 5'-end of the *cps* gene cluster are four relatively conserved open reading frames - *cpsA* (*wzg*)-*cpsB* (*wzh*)-*cpsC* (*wzd*)-*cpsD* (*wze*). Sequence differences in this region were used to classify 11 *S. pneumoniae* serotypes into two classes and, in the region between the 3'-end of *cpsA* and the 5'-end of *cpsB*, there were sites of heterogeneity between and within serotypes (Jiang et al., 2001; Lawrence et al., 2000). *S. pneumoniae* is characterised by high frequency recombination within the *cps* gene

cluster, leading to serotype "switching" among isolates within genetic lineages defined by relationships between their more conserved housekeeping genes (Coffey et al., 1998; Jiang et al., 2001).

Pneumococcal serogroup/type identification is currently performed, using large panels of expensive antisera, by various methods, including capsular swelling (Quellung) reaction - the traditional "gold standard"- latex agglutination and coagglutination (Arai et al., 2001; Lalitha et al., 1999). Cross-reactions between serotypes and discrepancies between methods can occur and some strains are nonserotypable (Henrichsen, 1999).

There is a need for further methods which can be used to identify different *Streptococcus pneumoniae* serotypes.

SUMMARY OF THE INVENTION

Through the complex analysis of a large number of polymorphisms which exist between 71 molecular capsular types (mct) and subtypes (mcst) of *Streptococcus pneumoniae* the present inventors have devised methods which can be used to distinguish between a significant number of different *S. pneumoniae* serotypes.

In a first aspect, the present invention provides a method of determining the serotype of *Streptococcus pneumoniae* in a sample, the method comprising analysing at least a portion of the nucleotide sequence between the 3' end of the *cpsA* gene and the 5' end of the *cpsB* gene.

In a preferred embodiment, the portion of the nucleotide sequence between the 3' end of the *cpsA* gene and the 5' end of the *cpsB* gene which is analysed is any nucleotide which is polymorphic between at least some of the *S. pneumoniae* serotypes referred to in Figure 2.

In a particularly preferred embodiment, the method comprises amplifying at least a portion of the nucleotide sequence between the 3' end of the *cpsA* gene and the 5' end of the *cpsB* gene, and sequencing the amplification product. More preferably, the entire approximate 800 bp region as provided in Figure 2 is amplified and sequenced.

In the case of sequencing to identify the serotype, the sequencing primers are selected such that they hybridise specifically to a region within or near to a region within which a polymorphism is present. The primers need not be specific to particular serotypes since it is the actual sequence information obtained during the sequencing process which is used to determine the *S. pneumoniae* serotype. Thus the primers may hybridise specifically to genomic DNA from all *S. pneumoniae* serotypes (or at least

those serotypes referred to in Figure 2), or to genomic DNA from some, but not all, *S. pneumoniae* serotypes.

When a portion of the nucleotide sequence between the 3' end of the *cpsA* gene and the 5' end of the *cpsB* gene is amplified, it is preferable that the amplification is performed using primer pairs comprising a sequence selected from the group consisting of:

1) GGCATT(/C)TATGGAGTTGATTCG(/A)TCCATT(/C)CACAC(C/T)TTAG and GC(/T)TCAATG(/A)TGG(/A)GCAATG(/T)ACTGGA(/C)GTA(/G)ATTCCCA(/G)ACATC,

2) GGCATT(/C)TATGGAGTTGATTCG(/A)TCCATT(/C)CACACC(/T)TTAG and CCATCAC(/T)ATAGAGGTTAC(/A)TG(/A)TCTGGCATT(/C)GC, and

3) GAAAGTGGG(/A/T)GGG(/A/T)A(/G)A(/C)T(/G)TAT(/C)AAAGTA(/G)AATTCT(/G)CAAGAT(/C)TTA(/G)AAA(/G)G and T(/G)CATG(/A)CTA(/G)AAC(/T)TCT(/A)ATC(/T)AAG(/A)GCATAACGACTATC(/T).

In an alternate embodiment, the nucleotide sequence analysis step comprises determining whether a polynucleotide obtained from *S. pneumoniae* selectively hybridises to a polynucleotide probe comprising one or more polymorphic regions of the nucleotide sequence between the 3' end of the *cpsA* gene and the 5' end of the *cpsB* gene, wherein such polymorphic regions are shown in Figure 2. More preferably, the nucleotide sequence analysis step comprises a plurality of said polynucleotide probes. In a particularly preferred embodiment, where hybridisation to a plurality of probes is used as a means of analysis, the plurality of polynucleotide probes are present as a microarray.

It has been noted that the method of the first aspect does not enable the identification of all known *S. pneumoniae* serotypes, for example shared sequences were noted in the following cases; 6A with 6B, 10A and 23A with 23F, 15B with 22F and 17F with 35B. Accordingly, in these instances further analysis will need to be performed to determine the correct serotype. To this end, the present inventors have discovered that polymorphisms in the *wzy* and/or *wzx* genes can also be useful for *S. pneumoniae* serotyping.

Accordingly, in a second aspect the present invention provides a method of determining the serotype of *Streptococcus pneumoniae* in a sample, the method comprising analysing at least a portion the *wzy* and/or *wzx* gene(s).

In a preferred embodiment, the method of the second aspect comprises amplifying at least a portion of the *wzy* and/or *wzx* gene(s), and determining the length of the amplification product.

In a particularly preferred embodiment, at least a portion of the *wzy* and/or *wzx* gene(s) is amplified using primer pairs comprising a sequence selected from the group consisting of:

1) GTAGGTGTAGTTTTTTCAGGGACTTTAATTTTATGCAGTG and TCGCTTAACACAATGGCTTTAGAAGGTAGAG,

2) GTTATTTTATTTTTTTGTCGGCATTGTATTCTTTATATCG and CAAATTCATCGTTTGTATCCATTTAACTGCATC,

3) CTTATATCTAATTATGTTCCGTCTATATTTATATGGGTTTGCTTTC and TTTCTCTTCATTTTCCTGATAATTTTGTACTTCTGAATG,

4) ATGCTTTTAAATTTCTTATTCATATCTATTTTTC and GTAAACAGAGAGCGAGTGATCATTTTAAAACTTTTGG,

5) G(/A)GATTTT(/G)TTTCAACCT(/C)GCAGTAATTTTAACAA(/C)TC(/T) G(/A) and CCTGAAAACAA(/G)TACT(/C)ACTTTCTGAATTTCAC(/T)GGA(/G)TATAAAG,

6) GTTTTATTGACTTTAAAGATGTTAGTTTCTTCGATTCCAG and TTTTTATTACTCTTCTTAAATCATAATGAATCGTACCAATCAAC,

7) GGATCAATGGCAACTATATTTACCCTACTCTCCACAG and GAGTCGAAACCAACCGGAAAAAGCAATTGAG,

8) CCTTTGGTTTATTATCCTACTTCCAAAACAGTTTATGC and CATATATCTCTTTATCCTGTCAATATTGATTGGCATTTTC,

9) GATATTAGCTATACCAACAATTGTTCTTTTCCTGTACTCAGTC and GCATTTCTAGTACCGAACCATTGAAACTATCATCTG,

10) GAAATTATAGTCGGAGCTTTCATTTATATTAGTTTACTGGTTCTG and CAGAATAAAGAGAGCTGTAATAGGTGCAACTTCATGC,

11) CTGTAATGTTTCTAATTAGTTCAGTATTTGCACTGGTTAATTC and CCCGTATATCCATTACTAAGAACAAGGTTGTATATTTCCTTC,

12) GTTTCTCATTAGTTCTGTATTTGCCCTTATTAATGTGC and CCATGGCTAAGTGCAAGATTATGAATCTCTCTC,

13) GTTTCTTATGTTTACCCTCAGCTTATATTGGCACAG and GATACCACAAATCTCCGAATTCTCTTAAAATAGATGG,

14) TTAAGTAGTTCACAAGTGATAGTGAACTTGGGATTGTC and CACTGAGATTATTTATTAGCTTTATCGGTAAGGTGGATAAG,

15) ATTACTTGTAATACTATGTATTCAACTAGTCA(/C)AGGATTTGAT

GG and GAACAAATTTCCGTATCAGATTTGCGATTTC,

16) CCAATGAAAAGGAAAGTTCAATGTGTTTTGTTTCTGC and GGTGCTTCAGCAAAAATCCCCGTATTTCTTATCAG,

17) TAGCTGATGTTCCGATAAATTATGGTGGGTAATAATAG and CTGCGACACTGTATATACCTACATTATAACTACTAGACATTTGC,

18) GCAACTTTGGTTCTAAAATTTTAGTCTTTTTAATGGTTCC and TGTTAAACCCCAATATAGAAATTGTATTGAGAATAGCAGC,

19) CGTTAATAGCTTATGTTCAACTGGTGATTGATTTTGG and TGATAGTTTTAGAAATAATATAAGGAATTGCAACTGCATGC,

20) TTCATGTC(/T)T(/C)TTTTG(/A)TCTAATCTGATTACAATTG(/C) TC(/T)A CAT CG(/A) and T(/C)GCATTTG(/T)GATCTGTCACAA(/G)TCAATAAGTTAAAACC,

21) GGTAGGTATTTTAATTGGAGGAAGAGAGTCTTGAATGG and ATCTTCCCTTCATAAATTGACATAGGAAAAATAAGAGCC,

22) CAATTCTAACTATGTCCAGTTTTATTTTTCCACTCATCAG and GACGTGATAATAATAAGCTGCCATTCCTGTCTAAAACG,

23) CGGCGGTATTAAGTAGAATATTAACACCTGAAGAGTATGGC and GGCAATCAGACTCAATAAGTTCATCCGTTTAAAGTTC,

24) GGTATTGCCTTTCCTTTGATAACTTCTCCTTATTTATCAC and TGAACTTGTAACTCGACACCCAAAAATATAAATAAATGAG,

25) GAATCGGACAATAGCACAGGTACGAACAAG and GCCATGTAATCAACTGACCAAGCAGGGTACTC,

26) CAAAGGAACGTTATCAGCAATTGTGTCAAATTTCAG and AAGATTAGGGCGCACAAAGTTTACTTGTTTTAGC,

27) GTTATTTCTTCAAATCTGCTCATAGTTTTAACCTCATCAC and TATCTTGCGTTTTCATCCCTTACAGTTATTAGGTTCAAAG,

28) TTCTTCAAATCTTTTGACAGTCTTGACCTCTTCCTTG and TATCGTGCATTCGAATCTGTTACAGCTAATACATTTAAAC,

29) GTCCTGACGCTATCAAATATCATTTTCCCATTAATCAC and CCCACATGTGATCAATAGGAGTGAAAATTCTCTATTC,

30) GCTTTGGCTAACTTTTCATCAAAGATTTTAATTTTTTTGTTAG and CCAGAGATAGCTGTAACACCAATTTTATCAATTCCCTTAG,

31) CCTTTGGCTAATTTCTTGGACGATAATGAATTTGTATATG and CCACAAACATTAGCAATAAAGAAACCTAACAATCCC,

32) GATCATACTCCCTATCATTACGACTCCCTATGTAACG and CCAAGAAATATCCAAACCTTTTGACACTAAACTTAATCC, and

33) GTTGTTTTAGCTCAAGGAGGGATAATGTTGGCTTCG and GCTGATTTTACAAATAGGAAAATAGAGATTGCACCAAC.

Guidance regarding the serotypes these primer pairs target, and the length of resulting amplification products, is provided in Tables 2 and 3.

It has been noted that some of the above primer pairs formed non-serotype specific amplicons, for example; PCR targeting serotype 6B also amplified 6A; PCR targeting 18C amplified all serotypes in serogroup 18; PCR targeting *wzx* (but not *wzy*) of serotype 23F, amplified three serotype 23A strains; PCR targeting *wzx* and *wzy* of serotypes 33/37 amplified a 33A isolate and that targeting *wzx* amplified a serotype 33B isolate. Accordingly, in these instances further analysis will need to be performed to determine the correct serotype. For instance, traditional serological typing can be performed.

As the skilled addressee would be aware, serotype 3 does not contain *wzy* and *wzx* genes. Accordingly, upon obtaining results using the methods of the first aspect, the presence of serotype 3 can be confirmed by analysing the orf2 (wze)-cap3A-cap3B region. Preferably, serotype 3 is identified by amplifying a portion of the orf2 (wze)-cap3A-cap3B region using primer pairs selected from the group consisting of:

1) GCACAAAAAAAAGTTTGATATTCCCCTTGACAATAG and GCAGGATCTAAGGAGGCTTCAAGATTCAACTC, and

2) CGAACCTACTATTGAGTGTGATACTTTTATGGGATACAGAG and CTGACAGCATGAAAATATATAACCGCCCAACGAATAAG.

During routine analysis of a sample comprising bacteria it will typically be desirable to ensure that the sample being analysed actually contains *Streptococcus pneumoniae*. Thus, it is preferred that the methods of the present invention include detecting any serotype of *Streptococcus pneumoniae* in the sample.

Such methods are known in the art and include, but are not limited to, amplifying portions of the psaA and/or pneumolysin genes followed by detection of the amplification products.

In a preferred embodiment, a portion of the psaA gene is amplified using primers comprising the sequence TACATTACTCGTTCTCTTTCTTTCTGCAATCATTCTTG and TAGTAGCTGTCGCCTTCTTTACCTTGTTCTGC. In another preferred embodiment, a portion of the pneumolysin gene is amplified using primers comprising the sequence AGAATAATCCCACTCTTCTTGCGGTTGA and CATGCTGTGAGCCGTTATTTTTCATACTG.

The present inventors have observed a strong correlation between the molecular typing techniques of the first and second aspect and the actual serotype of a strain as determined by traditional antibody based serological typing. However, the typing methods of the invention may be assisted by further serotyping the *S. pneumoniae* strain. For instance, to ensure recombination events have not occurred, upon typing with the methods of the invention the serotype can be confirmed by serologically typing for the strain suggested by the methods of the invention. Furthermore, the inventors have noted that a few serotypes are difficult to resolve using the methods of the invention. These serotypes include 6A and 6B; 10A, 23F and 23A; 15B and 22F; and 17F and 35B. Upon identification of any of these serotypes by the molecular techniques of the invention the serotype can be unequivocally typed using traditional serological methods.

In a third aspect, the present invention provides a polynucleotide comprising a sequence selected from those provided in Figures 2 to 64, or a fragment thereof which is at least 10 nucleotides in length, with the proviso the polynucleotide does not comprise the 3' end of the *cpsA* gene to the 5' end of the *cpsB* gene of a *S. pneumoniae* serotype selected from the group consisting of: 1, 2, 3, 4, 6A, 6B, 8, 9V, 14, 18C, 19F, 19A, 19B, 23F, 33F and 37, with the further proviso that the polynucleotide does not comprise the entire *wzy* and/or *wzx* gene(s) of a *S. pneumoniae* serotype selected from the group consisting of: 1, 2, 4, 6A, 6B, 8, 9V, 14, 18C, 19F, 19A, 19B, 23F, 33F and 37, or the entire *wzx* gene of *S. pneumoniae* serotype 19C.

In a preferred embodiment, the polynucleotide of the third aspect is at least 15 nucleotides, more preferably at least 20 nucleotides, more preferably at least 25 nucleotides, more preferably at least 30 nucleotides, more preferably at least 50 nucleotides in length, and even more preferably at least 100 nucleotides in length.

In a fourth aspect, the present invention provides a polynucleotide consisting essentially of 10 to 50 contiguous nucleotides corresponding to a portion of the 3' end of the *cpsA S. pneumoniae* gene or the 5' end of the *cpsB S. pneumoniae* gene, wherein said polynucleotide comprises one or more nucleotides which differ between different *S. pneumoniae* serotypes.

Polynucleotides of the fourth aspect can be used as amplification primers, or as probes, for the identification of different *S. pneumoniae* serotypes.

Preferably the nucleotides which differ between *S. pneumoniae* serotypes correspond to one or more of positions as shown in Figure 2.

Preferably, the polynucleotide of the fourth aspect is detectably labelled. The label can be any suitable label known in the art including, but not limited to, radionuclides, enzymes, fluorescent, and chemiluminescent labels.

In a fifth aspect, the present invention provides a polynucleotide consisting essentially of 10 to 50 contiguous nucleotides corresponding to a portion of the *S. pneumoniae wzy* and/or *wzx* gene(s), wherein said polynucleotide comprises one or more nucleotides which differ between different *S. pneumoniae* serotypes.

In a sixth aspect the present invention provides a composition comprising a plurality of polynucleotides according to the invention. Preferably, the composition further comprises a carrier or excipient. Preferably, the carrier or excipient is water or a suitable buffer. The composition may be used in methods of typing different *S. pneumoniae* serotypes.

In a seventh aspect the present invention provides a microarray comprising a plurality of polynucleotides according to the invention. The microarray may be used in methods of typing different *S. pneumoniae* serotypes.

In an eighth aspect, the present invention provides a kit comprising at least one polynucleotide of the present.

Preferably, the polynucleotide is in accordance with the fourth of fifth aspects of the invention. In one embodiment, the kit further comprises reagents necessary for nucleic acid amplification. In another embodiment, the polynucleotide of the fourth or fifth aspect are detectably labelled and the kit further comprises means for detecting the labelled polynucleotide.

As will be apparent, preferred features and characteristics of one aspect of the invention are applicable to many other aspects of the invention.

Throughout this specification the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

The invention is hereinafter described by way of the following non-limiting examples and with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE ACCOMPANYING DRAWINGS

Figure 1. The genomic sequence of *cpsA* (*wzg*) and *cpsB* (*wzh*) genes of serotype 4 of *S. pneumoniae* as published by Jiang et al. (2001) and deposited as GenBank Accession Number AF316639. The remaining 3' sequence of GenBank Accession Number

AF316639 has not been provided. Nucleotides 1520 to 2965 encode *cpsA* whilst nucleotides 2967 to 3698 encode *cpsB*.

Figure 2. Multiple sequence alignments for the region between the 3'-end of *cpsA (wzg)* and the 5'-end of *cpsB (wzh)* of 51 molecular capsular types (mct)/71 molecular capsular subtypes (mcst) of *S. pneumoniae*. The alignment numbering start point "1" refer to the position "2470" of *S. pneumoniae* serotype 4 *cpsA* (*wzg*) gene (GenBank accession number: AF316639) (Figure 1).

Figure 3: Partial sequence of strain 00-251-3185 of *S. pneumoniae wzx* gene.

Figure 4: Partial sequence of strain 01-122-0226 of *S. pneumoniae wzx* gene.

Figure 5: Partial sequence of strain 01-192-2471 of *S. pneumoniae wzx* gene.

Figure 6: Partial sequence of strain MA055100 of *S. pneumoniae wzx* gene.

Figure 7: Partial sequence of strain NZSPN01/329 of *S. pneumoniae wzx* gene.

Figure 8: Partial sequence of strain 00-256-1986 of *S. pneumoniae wzx* gene.

Figure 9: Partial sequence of strain NZSPN01/276 of *S. pneumoniae wzx* gene.

Figure 10: Partial sequence of strain 00-201-1422 of *S. pneumoniae wzx* gene.

Figure 11: Partial sequence of strain 00-211-1669 of *S. pneumoniae wzx* gene.

Figure 12: Partial sequence of strain 00S002 of *S. pneumoniae wzx* gene.

Figure 13: Partial sequence of strain 00-251-3185 of *S. pneumoniae wzy* gene.

Figure 14: Partial sequence of strain 01-122-0226 of *S. pneumoniae wzy* gene.

Figure 15: Partial sequence of strain 01-192-2471 of *S. pneumoniae wzy* gene.

Figure 16: Partial sequence of strain MA055100 of *S. pneumoniae wzy* gene.

Figure 17: Partial sequence of strain NZSPN01/329 of *S. pneumoniae wzy* gene.

Figure 18: Partial sequence of strain 00-256-1986 of *S. pneumoniae wzy* gene.

Figure 19: Partial sequence of strain NZSPN01/276 of *S. pneumoniae wzy* gene.

Figure 20: Partial sequence of strain 00-201-1422 of *S. pneumoniae wzy* gene.

Figure 21: Partial sequence of strain 00-211-1669 of *S. pneumoniae wzy* gene.

Figure 22: Partial sequence of strain 00S002 of *S. pneumoniae wzy* gene.

Figure 23: Partial sequence of strain NZSPN01/509 of *S. pneumoniae cpsI* and *wzx* genes.

Figure 24: Partial sequence of strain MA050408 of *S. pneumoniae cpsI* and *wzx* genes.

Figure 25: Partial sequence of strain MA052433 of *S. pneumoniae cpsI* and *wzx* genes.

Figure 26: Partial sequence of strain 00S009 of *S. pneumoniae cpsI* and *wzx* genes.

Figure 27: Partial sequence of strain 99-325-0373 of *S. pneumoniae cpsI* and *wzx* genes.

Figure 28: Partial sequence of strain NZSPN00/454 of *S. pneumoniae cpsI* and *wzx* genes.

Figure 29: Partial sequence of strain NZSPN00/484 of *S. pneumoniae cpsI* and *wzx* genes.

Figure 30: Partial sequence of strain 00-081-2291 of *S. pneumoniae wzy* and *wzx* genes.

Figure 31: Partial sequence of strain 00S168 of *S. pneumoniae wzy* and *wzx* genes.

Figure 32: Partial sequence of strain 00-280-1493 of *S. pneumoniae wzy* and *wzx* genes.

Figure 33: Partial sequence of strain MA063073 of *S. pneumoniae wzy* and *wzx* genes.

Figure 34: Partial sequence of strain NZSPN00/410 of *S. pneumoniae wzy* and *wzx* genes.

Figure 35: Partial sequence of strain NZSPN01/243 of *S. pneumoniae wzy* and *wzx* genes.

Figure 36: Partial sequence of strain MA063087 of *S. pneumoniae wzy* and *wzx* genes.

Figure 37: Partial sequence of strain MA063207 of *S. pneumoniae wzy* and *wzx* genes.

Figure 38: Partial sequence of strain 01S333 of *S. pneumoniae wzx* gene.

Figure 39: Partial sequence of strain MA050663 of *S. pneumoniae wciW* and *wzx* genes.

Figure 40: Partial sequence of strain 01S319 of *S. pneumoniae wciW* and *wzx* genes.

Figure 41: Partial sequence of strain NZSPN00/353 of *S. pneumoniae wciW* and *wzx* genes.

Figure 42: Partial sequence of strain MA062610 of *S. pneumoniae wciW* and *wzx* genes.

Figure 43: Partial sequence of strain MA053392 of *S. pneumoniae wciW* and *wzx* genes.

Figure 44: Partial sequence of strain NZSPN00/319 of *S. pneumoniae wciW* and *wzx* genes.

Figure 45: Partial sequence of strain NZSPN01/278 of *S. pneumoniae wciW* and *wzx* genes.

Figure 46: Partial sequence of strain 01S009 of *S. pneumoniae wciW* and *wzx* genes.

Figure 47: Partial sequence of strain MA052628 of *S. pneumoniae wciW* and *wzx* genes.

Figure 48: Partial sequence of strain 00-081-2291 of *S. pneumoniae cpsJ* and *wzy* genes.

Figure 49: Partial sequence of strain 00-280-1493 of *S. pneumoniae cpsJ* and *wzy* genes.

Figure 50: Partial sequence of strain NZSPN00/410 of *S. pneumoniae cpsJ* and *wzy* genes.

Figure 51: Partial sequence of strain NZSPN01/243 of *S. pneumoniae cpsJ* and *wzy* genes.

Figure 52: Partial sequence of strain MA063073 of *S. pneumoniae cpsJ* and *wzy* genes.

Figure 53: Partial sequence of strain 00S168 of *S. pneumoniae cpsJ* and *wzy* genes.

Figure 54: Partial sequence of strain MA063087 of *S. pneumoniae cpsJ* and *wzy* genes.

Figure 55: Partial sequence of strain MA063207 of *S. pneumoniae cpsJ* and *wzy* genes.

Figure 56: Partial sequence of strain 01S319 of *S. pneumoniae wzx* and *wzy* genes.

Figure 57: Partial sequence of strain NZSPN00/353 of *S. pneumoniae wzx* and *wzy* genes.

Figure 58: Partial sequence of strain MA062610 of *S. pneumoniae wzx* and *wzy* genes.

Figure 59: Partial sequence of strain MA053392 of *S. pneumoniae wzx* and *wzy* genes.

Figure 60: Partial sequence of strain NZSPN00/319 of *S. pneumoniae wzx* and *wzy* genes.

Figure 61: Partial sequence of strain NZSPN01/278 of *S. pneumoniae wzx* and *wzy* genes.

Figure 62: Partial sequence of strain MA050663 of *S. pneumoniae wzx* and *wzy* genes.

Figure 63: Partial sequence of strain MA052628 of *S. pneumoniae wzx* and *wzy* genes.

Figure 64: Partial sequence of strain 01S009 of *S. pneumoniae wzx* and *wzy* genes.

Figure 65: Phylogenetic tree inferred from sequences in the region between the3'-end of *cpsA (wzg)* and the 5'-end of *cpsB (wzh)* genes for 51 molecular capsular subtypes (mct)/71 molecular capsular subtypes (mcst) of *S. pneumoniae.* Most of the tree input sequences are from Figure 2 and Table 1; for GenBank accession numbers see Table 1. Sequences of two nonserotypable isolates were also included; they were clearly separated from the other known mct/mcst.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art (e.g., in cell culture, molecular genetics, nucleic acid chemistry, hybridization techniques and biochemistry).

As used herein, the term "nucleotide sequence between the 3' end of the *cpsA* gene and the 5' end of the *cpsB* gene" at least refers to the region spanning from nucleotide 2470 to nucleotide 3268 of Figure 1. Figure 1 provides the genomic sequence of *cpsA* (*wzg*) and *cpsB* (*wzh*) genes of serotype 4 as published by Jiang et al. (2001) and submitted as GenBank Accession Number AF316639. As the skilled addressee would be aware, the same region from other serotypes of *S. pneumoniae* can be identified using standard techniques such as DNA cloning, sequencing and nucleotide sequence alignment. Such techniques are described in further detail in the Examples section. In addition, these techniques have been used to determine the nucleotide sequence between the 3' end of the *cpsA* gene and the 5' end of the *cpsB* gene from many different serotypes of *S. pneumoniae*, the results of which, including a consensus sequence for this region, are also provided in Figure 2.

General Techniques

Unless otherwise indicated, the recombinant DNA and immunological techniques utilized in the present invention are standard procedures, well known to those skilled in the art. Such techniques are described and explained throughout the literature in sources such as, J. Perbal, A Practical Guide to Molecular Cloning, John Wiley and Sons (1984), J. Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory Press (1989), T.A. Brown (editor), Essential Molecular Biology: A Practical Approach, Volumes 1 and 2, IRL Press (1991), D.M. Glover and B.D. Hames (editors), DNA Cloning: A Practical Approach, Volumes 1-4, IRL Press (1995 and 1996), and F.M. Ausubel et al. (editors), Current Protocols in Molecular Biology, Greene Pub. Associates and Wiley-Interscience (1988, including all updates until present), Ed Harlow and David Lane (editors) Antibodies: A Laboratory Manual, Cold Spring Harbour Laboratory, (1988), and J.E. Coligan et al. (editors) Current Protocols in Immunology, John Wiley & Sons (including all updates until present), and are incorporated herein by reference.

Detection of Polymorphisms

Any technique known in the art can be used to detect a polymorphism described herein. Examples of such techniques include, but are not limited to, sequencing of the DNA at one or more of the relevant positions; differential hybridisation of an oligonucleotide probe designed to hybridise at the relevant positions of a particular *S. pneumoniae* serotype(s); denaturing gel electrophoresis following digestion with an appropriate restriction enzyme, preferably following amplification of the relevant DNA

regions; S1 nuclease sequence analysis; non-denaturing gel electrophoresis, preferably following amplification of the relevant DNA regions; conventional RFLP (restriction fragment length polymorphism) assays; selective DNA amplification using oligonucleotides which are matched for a particular *S. pneumoniae* serotype(s) unmatched for other *S. pneumoniae* serotype(s); or the selective introduction of a restriction site using a PCR (or similar) primer matched for a particular *S. pneumoniae* serotype(s), followed by a restriction digest. As outlined above, it is preferred that the nucleotide sequence between the 3' end of the *cpsA* gene and the 5' end of the *cpsB* gene is characterized by DNA sequencing, whilst the analysis.at least a portion the *wzy* and/or *wzx* gene is performed by procedures involving the detection of amplification products.

PCR-based methods of detection may rely upon the use of primer pairs, at least one of which binds specifically to a region of interest in one or more, but not all, serotypes. Unless both primers bind, no PCR product will be obtained. Consequently, the presence or absence of a specific PCR product may be used to determine the presence of a sequence indicative of a specific *S. pneumoniae* serotype(s). However, as mentioned, only one primer need correspond to a region of heterogeneity in the genes/regions of interest. The other primer may bind to a conserved or heterogenous region within said gene/region or even a region within another part of the *S. pneumoniae* genome, whether said region is conserved or heterogeneous between serotypes.

Alternatively, primers that bind to conserved regions of the *S. pneumoniae* genome but which flank a region whose length varies between serotypes may be used. In this case, a PCR product will always be obtained when *S. pneumoniae* bacteria are present but the size of the PCR product varies between serotypes. Examples of such varying amplification product lengths are disclosed herein in relation to the *wzy* and *wzx* genes.

Furthermore, a combination of specific binding of one or both primers and variations in the length of PCR primer may be used as a means of identifying particular molecular serotypes.

In some cases, PCR and other specific hybridisation- based serotyping methods will involve the use of nucleotide primers/probes which bind specifically to a region of the genome of a *S. pneumoniae* serotype which includes a nucleotide which varies between two or more serotypes. Thus the primers/probes may comprise a sequence which is complementary to one of such regions. Where positions of heterogeneity are close together (for instance within 5 or so nucleotides), it may be desirable to use a

primer/probe which hybridises specifically to a region of the *S. pneumoniae* genome that comprises two or more positions of heterogeneity. Such primers/probes are likely to have improved specificity and reduce the likelihood of false positives.

PCR techniques that utilize fluorescent dyes may be used in the detection methods of the present invention. These include, but are not limited to, the following five techniques.

i) Fluorescent dyes can be used to detect specific PCR amplified double stranded DNA product (e.g. ethidium bromide, or SYBR Green I).

ii) The 5' nuclease (TaqMan) assay can be used which utilizes a specially constructed primer whose fluorescence is quenched until it is released by the nuclease activity of the Taq DNA polymerase during extension of the PCR product.

iii) Assays based on Molecular Beacon technology can be used which rely on a specially constructed oligonucleotide that when self-hybridized quenches fluorescence (fluorescent dye and quencher molecule are adjacent). Upon hybridization to a specific amplified PCR product, fluorescence is increased due to separation of the quencher from the fluorescent molecule.

iv) Assays based on Amplifluor (Intergen) technology can be used which utilize specially prepared primers, where again fluorescence is quenched due to self-hybridization. In this case, fluorescence is released during PCR amplification by extension through the primer sequence, which results in the separation of fluorescent and quencher molecules.

v) Assays that rely on an increase in fluorescence resonance energy transfer can be used which utilize two specially designed adjacent primers, which have different fluorochromes on their ends. When these primers anneal to a specific PCR amplified product, the two fluorochromes are brought together. The excitation of one fluorochrome results in an increase in fluorescence of the other fluorochrome.

Probes and primers may be fragments of DNA isolated from nature or may be synthetic. In one embodiment, primers/probes have a high melting temperature of >70°C so that they may be used in rapid cycle PCR. Preferably, the primers/probes comprise at least 10, 15 or 20 nucleotides. Typically, primers/probes consist of fewer than 50 or 30 nucleotides. Primers/probes are generally polynucleotides comprising deoxynucleotides. They may also be polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood

that the polynucleotides described herein may be modified by any method available in the art. Primers/probes may be labelled with any suitable detectable label such as radioactive atoms, fluorescent molecules or biotin.

The primers be synthesized using techniques which are well known in the art. Generally, the primers can be made using synthesizing machines which are commercially available.

If required, in order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme sites appended to their 5' ends. Thus, all nucleotides of the primers are derived from the gene sequence of interest or sequences adjacent to that gene except the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art.

A sample to be typed for the presence and/or identification of a *S. pneumoniae* serotype may be from a bacterial culture or a clinical sample from a patient, typically a human patient. Clinical samples may be cultured to produce a bacterial culture. However, it is also possible to test clinical samples directly with a culturing step.

The methods of the present invention can be used in a multi-step serotyping strategy. An example of such a multi-step serotyping strategy (algorithm) is shown in Table 6. However, a variety of other strategies are envisaged and can be designed by the skilled person using the sequence heterogeneity information presented herein. In particular, it is preferred that the serotyping procedure comprise at least one analysis step based on analysing one or regions between the 3' end of the *cpsA* gene and the 5' end of the *cpsB* gene. This analysis may optionally be combined with an analysis of one or more regions within the *wzy* and/or *wzx* genes.

<u>Microarrays</u>

Analysis of *S. pneumoniae* genomic sequences using the above techniques may take place in solution followed by standard resolution using methods such as gel electrophoresis. However in a preferred aspect of the invention, the primers/probes are immobilised onto a solid substrate to form arrays.

The polynucleotide probes are typically immobilised onto or in discrete regions of a solid substrate. The substrate may be porous to allow immobilisation within the substrate or substantially non-porous, in which case the probes are typically immobilised on the surface of the substrate. Examples of suitable solid substrates include flat glass (such as borosilicate glass), silicon wafers, mica, ceramics and organic polymers such as plastics, including polystyrene and polymethacrylate. It may also be possible to use semi-permeable membranes such as nitrocellulose or nylon

membranes, which are widely available. The semi-permeable membranes may be mounted on a more robust solid surface such as glass. The surfaces may optionally be coated with a layer of metal, such as gold, platinum or other transition metal.

Preferably, the solid substrate is generally a material having a rigid or semi-rigid surface. In preferred embodiments, at least one surface of the substrate will be substantially flat, although in some embodiments it may be desirable to physically separate synthesis regions for different polymers with, for example, raised regions or etched trenches. It is also preferred that the solid substrate is suitable for the high density application of DNA sequences in discrete areas of typically from 50 to 100 µm, giving a density of 10000 to 40000 cm^{-2}.

The solid substrate is conveniently divided up into sections. This may be achieved by techniques such as photoetching, or by the application of hydrophobic inks, for example teflon-based inks (Cel-line, USA). Discrete positions, in which each different probes are located may have any convenient shape, e.g., circular, rectangular, elliptical, wedge-shaped, etc.

Attachment of the library sequences to the substrate may be by covalent or non-covalent means. The library sequences may be attached to the substrate via a layer of molecules to which the library sequences bind. For example, the probes may be labelled with biotin and the substrate coated with avidin and/or streptavidin. A convenient feature of using biotinylated probes is that the efficiency of coupling to the solid substrate can be determined easily. Since the polynucleotide probes may bind only poorly to some solid substrates, it is often necessary to provide a chemical interface between the solid substrate (such as in the case of glass) and the probes. Thus, the surface of the substrate may be prepared by, for example, coating with a chemical that increases or decreases the hydrophobicity or coating with a chemical that allows covalent linkage of the polynucleotide probes. Some chemical coatings may both alter the hydrophobicity and allow covalent linkage. Hydrophobicity on a solid substrate may readily be increased by silane treatment or other treatments known in the art. Examples of suitable chemical coatings include polylysine and poly(ethyleneimine). Further details of methods for the attachment of are provided in US 6,248,521.

Techniques for producing immobilised arrays of nucleic acid molecules have been described in the art. A useful review is provided in Schena *et al.* (1998), which also gives references for the techniques described therein.

Microarray-manufacturing technologies fall into two main categories—synthesis and delivery. In the synthesis approaches, microarrays are prepared in a stepwise

fashion by the *in situ* synthesis of nucleic acids from biochemical building blocks. With each round of synthesis, nucleotides are added to growing chains until the desired length is achieved. A number of prior art methods describe how to synthesise single-stranded nucleic acid molecule libraries *in situ*, using for example masking techniques (photolithography) to build up various permutations of sequences at the various discrete positions on the solid substrate. US 5,837,832 describes an improved method for producing DNA arrays immobilised to silicon substrates based on very large scale integration technology. In particular, U.S. Patent No. 5,837,832 describes a strategy called "tiling" to synthesize specific sets of probes at spatially-defined locations on a substrate which may be used to produced the immobilised DNA libraries of the present invention. US 5,837,832 also provides references for earlier techniques that may also be used.

The delivery technologies, by contrast, use the exogenous deposition of prepared biochemical substances for chip fabrication. For example, DNA may also be printed directly onto the substrate using for example robotic devices equipped with either pins (mechanical microspotting) or piezo electric devices (ink jetting). In mechanical microspotting, a biochemical sample is loaded into a spotting pin by capillary action, and a small volume is transferred to a solid surface by physical contact between the pin and the solid substrate. After the first spotting cycle, the pin is washed and a second sample is loaded and deposited to an adjacent address. Robotic control systems and multiplexed printheads allow automated microarray fabrication. Ink jetting involves loading a biochemical sample, such as a polynucleotide into a miniature nozzle equipped with a piezoelectric fitting and an electrical current is used to expel a precise amount of liquid from the jet onto the substrate. After the first jetting step, the jet is washed and a second sample is loaded and deposited to an adjacent address. A repeated series of cycles with multiple jets enables rapid microarray production.

In one embodiment, the microarray is a high density array, comprising greater than about 50, preferably greater than about 100 or 200 different nucleic acid probes. Such high density probes comprise a probe density of greater than about 50, preferably greater than about 500, more preferably greater than about 1,000, most preferably greater than about 2,000 different nucleic acid probes per cm^2. The array may further comprise mismatch control probes and/or reference probes (such as positive controls).

Microarrays of the invention will typically comprise a plurality of primers/probes as described above. The primers/probes may be grouped on the array in any order.

Elements in an array may contain only one type of probe/primer or a number of different probes/primers.

Detection of binding of *S. pneumoniae* DNA to immobilised probes/primers may be performed using a number of techniques. For example, the immobilised probes which are specific for one or a number of serotypes, may function as capture probes. Following binding of the genomic DNA to the array, the array is washed and incubated with one or more labelled detection probes which hybridise specifically to regions of the *S. pneumoniae* genome which are conserved (for example the *S. pneumoniae* psaA or pneumolysin probes/primers described herein could be utilized for this purpose). The binding of these detection probes may then be determined by detecting the presence of the label. For example, the label may be a fluorescent label and the array may be placed in an X-Y reader under a charge-coupled device (CCD) camera.

Other techniques include labelling the genomic DNA prior to contact with the array (using nick-translation and labelled dNTPs for example). Binding of the genomic DNA can then be detected directly.

It is also possible to employ a single PCR amplification step using labelled dNTPs. In this embodiment, the genomic DNA fragment binds to a first primer present in the array. The addition of polymerase, dNTPs, including some labelled dNTPs and a second primer results in synthesis of a PCR product incorporating labelled nucleotides. The labelled PCR fragment captured on the plate may then be detected.

A number of available detection techniques do not require labels but instead rely on changes in mass upon ligand binding (e.g. surface plasmon resonance- SPR). The principles of SPR and the types of solid substrates required for use in SPR (e.g. BIACore chips) are described in Ausubel *et al.*, Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc.

Examples of the utilization of microarrays in genotyping include the use of microarrays to differentiate between closely related *Cryptosporidium parvum* isolates and *Cryptosporidium* species (Straub et al., 2002), and the use of microarrays to differentiate between species of *Listeria* (Volokhov et al., 2002). The detection principles applied in these studies can be used with the polymorphisms/primers/probes identified by the present inventors to identify different serotypes of *S. pneumoniae* in a sample.

<u>Kits</u>

In one embodiment, kits of the present invention include, in an amount sufficient for at least one assay, a polynucleotide probe of the invention which

preferentially hybridizes to a target nucleic acid sequence in a test sample under hybridization assay conditions. Kits containing multiple probes are also contemplated by the present invention where the multiple probes are designed to target different nucleic acid sequences from different *S. pneumoniae* serotypes and may include distinct labels which permit the probes to be differentially detected in a test sample. Kits according to the present invention may further comprise at least one of the following: (i) one or more amplification primers for amplifying a target sequence contained in or derived from the target nucleic acid; (ii) a capture probe for isolating and purifying target nucleic acid present in a test sample; and (iii) if a capture probe is included, a solid support material (e.g., magnetically responsive particles) for immobilizing the capture probe, either directly or indirectly, in a test sample. Kits of the present invention may further include one or more helper probes.

Typically, the kits will also include instructions recorded in a tangible form (e.g., contained on paper or an electronic medium) for using the packaged polynucleotide in a detection assay for determining the presence or amount of a target nucleic acid sequence in a test sample. The assay described in the written instructions may include steps for isolating and purifying the target nucleic acid prior to detection with the polynucleotide probe, and/or amplifying a target sequence contained in the target nucleic acid. The instructions will typically indicate the reagents and/or concentrations of reagents and at least one assay method parameter which might be, for example, the relative amounts of reagents to use per amount of sample. In addition, such specifics as maintenance, time periods, temperature and buffer conditions may also be included.

Uses

As discussed above, *S. pneumoniae* is a leading cause of morbidity and mortality causing invasive disease such as meningitis and pneumonia as well as more localised disease such as acute otitis media and sinusitis. Continued surveillance is critical to monitor vaccine efficacy and changes in incidence and distribution of colonising and invasive serotypes. Any increase in disease caused by previously uncommon nonvaccine serotypes could necessitate a change in vaccine composition. Thus, the detection methods, probes/primer and microarrays of the invention may be used to monitor the epidemiology of invasive *S. pneumoniae* infections to assist in disease control and to inform vaccine policy.

The molecular typing methods of the invention may also assist in comprehensive serotype identification that will be useful for epidemiological and other

related studies that will be needed to monitor *S. pneumoniae* before and after introduction of *S. pneumoniae* vaccines.

Examples

MATERIALS AND METHODS

Pneumococcal reference panels (Table 1)

Reference panels 1-4, which consisted of 118 isolates, were kindly provided and serotyped by colleagues in Australia and Canada. All had been serotyped using the standard Quellung method and included all 23 serotypes represented in the polysaccharide vaccine, and 28 additional serotypes; there were multiple isolates of 40 serotypes and five isolates that could not be serotyped with available antisera. Reference panel 5 consisted of 21 invasive isolates from our diagnostic laboratory at the Centre for Infectious Diseases and Microbiology (CIDM), Sydney, for which serotypes were known at the beginning of the study. These five reference panels were used for the development and preliminary evaluation of MCT methods. Panels 2 and 4 were tested by MCT, initially, without knowledge of the conventional serotyping (CS) results.

Table 1. Conventional serotyping (CS) and molecular capsular typing (MCT) results of *S. pneumoniae* strains used in this study.

Strain numbers and geographic origin	CS[1]	MCT-Seq/mctsp[2]	MCT-PCR[2]	GenBank[2] accession numbers
Reference panel 1[3]				
Queensland				
00S001	19F	19F	19F	AF532666
00S002	6B	6B-q	6B	AF532705; AY163180, AY163190
00S006	19A	19A	19A	AF532663
00S009	23F	23F-g	23F	AF532677; AY163214, AY163232
00S014	1	1	1	AF532632
00S016	9V	9V	9V	AF532710
00S023	5	5-q		AF532697
00S033	17F	17F-35B		AF532657
00S036	11A	11A-q		AF532637
00S042	18C	18C/18B	18C	AF532661
00S059	9N	9N		AF532709
00S063	12F	12F		AF532640
00S067	8	8	8	AF532708
00S124	7F	7F		AF532707
00S154	15B	15B-q		AF532649
00S159	4	4	4	

00S168	33F	33F-q	33F/37	AF532687; AY163199, AY163221
00S246	22F	22F		AF532673
00S259	2	2-q	2	AF532669
00S300	22A	22A		AF532672
01S009	18C	18C/18B	18C	
01S020	7C	7C		AF532706
01S043	10A	10A-q		AF532633
01S143	3	3	3	AF532682
01S146	10F	10F		AF532635
01S305	20	20/13		AF532670
01S319	18A	18A	18C	AF532658; AY163208, AY163224
01S333	33B	33B	33F-X; 33F-Y-NEG	AF532686
01S358	35B	35B		AF532691
01S666	14	14-g	14	AF532643
01S682	16F	16F		AF532653
01S691	15C	15C-q		AF532651
01S753	4	4	4	AF532693
Reference panel 2[4]				
Victoria				
0013856	35B	35B		
0013976	6A	6A-ca	6B	
0017666	9V	9V	9V	
0019532	23F	23F-g	23F	
0102206	8	8	8	
0103678	19F	19F	19F	
0104603	6B	6B-q	6B	
0104604	22F	22F		
0104912	4	4	4	
0105015	14	14-g	14	AF532644
Reference panel 3[5]				
Canada				
MA007753	31	31/42		AF532684
MA007765	5	5-q		
MA008229	10F	10F		AF532636
MA008562	11A	11A-q		
MA008622	31	31/42		
MA050408	23A	23A-23F	23F-X; 23F-Y-NEG	AF532674
MA050663	18F	18F	18C	AF532662; AY163207, AY163230
MA050910	2	2-q	2	
MA050947	38	38/25F		AF532712
MA051117	22A	22A		
MA051617	35F	35F		AF532692
MA051950	42	42/31		AF532695
MA052002	15A	15A-ca1		AF532646
MA052150	11B	11B		AF532639
MA052217	7C	7C		
MA052253	17F	17F-35B		
MA052433	23A	23A-ca	23F-X; 23F-Y-NEG	AF532675
MA052434	15A	15A-ca2		AF532647

MA052628	18C	18C/18B	18C	-; AY163215, AY163231
MA052979	15C	15C-ca		AF532652
MA053096	20	20/13		
MA053188	15B	15B-q		
MA053392	18B	18B/18C	18C	AF532660; AY163211, AY163227
MA053567	12F	12F		
MA053684	38	38/25F		
MA053782	13	13/20		AF532642
MA053909	35B	35B		
MA054004	13	13/20		
MA054006	13	13/20		
MA054242	38	38/25F		
MA054294	16F	16F		
MA054338	35F	35F		
MA054357	1	1	1	
MA054490	34	34		AF532690
MA054545	3	3	3	
MA054735	10A	10A-q		
MA054832	34	34		
MA054883	7F	7F		
MA055006	9V	9V	9V	
MA055054	22F	22F		
MA055100	6A	6A-ca	6B	AF532702; AY163174, AY163184
MA056382	19A	19A	19A	AF532664
MA059287	25F	25F/38		AF532711
MA061296	41F	41F		AF532694
MA061378	17A	17A		AF532655
MA061938	21	21		AF532671
MA062028	29	29		AF532680
MA062610	18B	18B/18C	18C	-; AY163210, AY163226
MA063013	9N	9N		
MA063073	33F	33F-g/33A	33F/37	AF532689; AY163201, AY163220
MA063087	33A	33A/33F-g	33F/37	AF532685; AY163204, AY163222
MA063189	Nonserotypeable	No-amplicon		
MA063207	37	37	33F/37	AF532713; AY163205, AY163223
MA063745	Nonserotypeable	Nonserotypeable-ca		AF532715
Reference panel 4[6]				
New South Wales				
00-177-0145	19A	19A	19A	
01-184-0091	18C	18C/18B	18C	
00-237-0230	17F	17F-35B		AF532656
01-273-0175	16F	16F		
00-201-0306	14	14-g	14	
01-117-0176	13	13/20		
01-239-0283	12F	12F		
00-206-0233	11A	11A-q		
00-222-0342	10A	10A-23F	23F-NEG	AF532634
01-180-0149	1	1	1	

01-122-0226	6A	6A-ca	6B	AF532698; AY163172, AY163182
99-308-0385	4	4		
00-234-0199	38	38/25F		
00-074-0065	35F	35F		
00-280-0121	3	3	3	
99-308-0290	23F	23F-g	23F	
00-244-0101	22F	22F		
00-250-0302	22A	22A		
00-244-0108	20	20/13		
01-009-0101	19F	19F	19F	AF532668
01-254-0150	7F	7F		
Reference panel 5[7] New South Wales, (CIDM)				
00-163-0650	14	14-g	14	
00-141-1399	19F	19F	19F	
00-070-0212	23F	23F-g	23F	
01-018-1842	4	4	4	
00-201-1422	6B	6B-g	6B	AF532703; AY163178, AY163188
00-180-2749	9V	9V	9V	
00-339-3084	9N	9N		
00-017-0985	11A	11A-q		
01-072-0391	12F	12F		AF532641
00-315-3100	15B	15B-c		AF532648
99-259-1456	18C	18C/18B	18C	
00-273-2862	4	4	4	
00-081-2291	33F	33F-g/33A	33F/37	-; AY163198, AY163216
00-118-2067	5	5-c		AF532696
01-175-0822	7F	7F		
00-324-0978	8	8	8	
00-152-1664	22F	22F		
00-211-1414	22F	22F		
00-200-0078	14	14-g	14	
00-118-0159	19F	19F	19F	
00-310-1104	4	4	4	
Clinical isolates New South Wales, (CIDM)[8]				
01-192-3558	6B	6B-g	6B	
01-192-2471	6A	6A-c	6B	AF532699; AY163173, AY163183
01-192-1205	6B	6B-g	6B	
01-191-1265	14	14-g	14	
01-189-0296	19F	19F	19F	
01-185-0511	15B	15B-22F		AF532650
01-184-0328	8	8	8	
01-179-2448	14	14-g	14	
01-178-0165	14	14-g	14	
01-176-3302	1	1	1	
01-173-2782	4	4	4	
01-170-0873	9V	9V	9V	
01-159-0505	14	14-g	14	

01-157-3399	4	4	4	
01-157-3394	4	4	4	
01-157-2062	4	4	4	
01-152-3295	14	14-g	14	
01-150-3706	14	14-g	14	
01-144-1862	7F	7F		
01-143-3353	4	4	4	
01-124-2300	12F	12F		
01-117-1910	4	4	4	
01-096-2050a	9V	9V	9V	
01-096-2050b	9V	9V	9V	
01-096-2027	9V	9V	9V	
01-077-1533	7F	7F		
01-075-3257	9N	9N		
01-058-3662	14	14-g	14	
01-048-1320	19A	19A	19A	
01-005-0764	19F	19F	19F	AF532650
00-361-1217	6B	6B-q	6B	
00-357-1164	14	14-g	14	
00-339-2918	9N	9N		
00-324-0977	8	8	8	
00-315-2993	23F	23F-g= 10A-23F	23F	
00-315-2254	23F	23F-g= 10A-23F	23F	
00-310-0630	14	14-g	14	
00-303-0303	19F	19F	19F	
00-293-1660	19F	19F	19F	
00-280-1493	33F	33F-q	33F/37	-; AY163200, AY163217
00-267-0653	8	8	8	
00-258-1120	14	14-g	14	
00-257-0881	9V	9V	9V	
00-256-1986	6A	6A-ca	6B	-; AY163176, AY163186
00-251-3185	6A	6A-6B-g= 6B-g	6B	AF532700; AY163171, AY163181
00-245-3950	23F	23F-g= 10A-23F	23F	
00-243-2229	3	3	3	
00-242-0394	14	14-g	14	
00-241-2964	9V	9V	9V	
00-238-3448	23F	23F-g= 10A-23F	23F	
00-235-3584	19F	19F	19F	AF532665
00-228-3777	35B	35B		
00-225-1482	3	3	3	
00-225-0333	19F	19F	19F	
00-217-3003	4	4	4	
00-211-1669	6B	6B-c	6B	AF532704; AY163179, AY163189
00-211-0475	22F	22F		
00-211-0469	22F	22F		
00-209-3409	3	3	3	
00-208-0179	4	4	4	

00-200-1013	14	14-g	14	
00-200-1012	14	14-g	14	
00-199-0498	4	4	4	
00-196-2923	9V	9V	9V	
00-192-2087	19A	19A	19A	
00-184-1203	6B	6B-q	6B	
00-181-1568	23F	23F-g= 10A-23F	23F	
00-181-1567	23F	23F-g= 10A-23F	23F	
00-173-3686	4	4	4	
00-164-1705	6B	6B-q	6B	
00-163-1533	14	14-g	14	
00-149-1265	7F	7F		
00-149-1264	7F	7F		
00-143-1473	15B	15B-22F		
00-138-3435	3	3	3	
00-118-2891	19F	19F	19F	
00-093-1315	3	3	3	AF532681
00-078-0883	14	14-g	14	
00-074-3370	14	14-g	14	
00-070-0212	23F	23F-g= 10A-23F	23F	
00-066-3506	4	4	4	
00-043-0876	19A	19A	19A	
00-036-1378	19F	19F	19F	
00-008-0865	8	8	8	
99-348-3354	6A	6A-ca	6B	
99-338-1052	19F	19F	19F	
99-325-0373	23F	23F-c	23F	AF532678
99-324-1010	4	4	4	
99-404-0191	4	4	4	
99-310-0070	4	4	4	
99-302-1894	9V	9V	9V	
99-293-1704	19A	19A	19A	
99-287-2376	35B	35B		
99-287-2320	35B	35B		
99-287-2298	35B	35B		
99-284-1034	14	14-c	14	AF532645
99-276-0568	9V	9V	9V	
99-242-0442A	6B	6B-q	6B	
99-241-1187A	4	4	4	
99-237-2839	9V	9V	9V	
99-235-2193	4	4	4	
99-226-1026B	7F	7F		
99-221-2755	9V	9V	9V	
99-221-2745A1	23F	23F-g= 10A-23F	23F	
99-221-0278	4	4	4	
99-218-2527	23F	23F-g= 10A-23F	23F	
99-201-1708	3	3	3	
99-196-2909B	10A	10A-23F =23F-g	23F-NEG	

99-196-2908B	10A	10A-23F= 23F-g	23F-NEG	
99-196-2882A	10A	10A-23F =23F-g	23F-NEG	
99-196-2880A	10A	10A-23F =23F-g	23F-NEG	
99-195-0430	14	14-g	14	
99-193-2919A	4	4	4	
99-193-2918B	4	4	4	
99-193-2747B	4	4	4	
99-193-2491A	18C	18C/18B	18C	
99-192-0047B	23F	23F-g= 10A-23F	23F	
99-188-2369A	4	4	4	
99-186-2831	7F	7F		
99-186-1038	14	14-g	14	
99-186-0417	14	14-g	14	
99-184-0894	14	14-g	14	
99-182-1919	4	4	4	
99-180-2653	4	4	4	
99-178-0901	14	14-g	14	
99-177-1060	11A	11A-q		
99-176-1983	18C	18C/18B	18C	
99-173-2956	4	4	4	
99-169-0432	6B	6B-g	6B	
99-159-2018	7F	7F		
99-158-1250	14	14-g	14	
99-157-0650	19F	19F	19F	
99-146-2324	19F	19F	19F	
99-144-1497	22F	22F		
99-134-2273	3	3	3	
99-132-2724	15B	15B-q		
99-132-2558	15B	15B-q		
99-132-2557	15B	15B-q		
99-130-2037	14	14-g	14	
99-110-2820	9N	9N		
99-108-0976	23F	23F-g= 10A-23F	23F	
99-107-0715	14	14-g	14	
99-104-1860	4	4	4	
99-099-0423	19F	19F	19F	
99-095-1044	20	20/13		
99-091-2295	23B	23B	23F-NEG	AF532676
99-090-2551	14	14-g	14	
99-090-2390	3	3	3	
99-090-2387	3	3	3	
99-033-2630	23F	23F-g= 10A-23F	23F	
99-028-0057	7C	7C		
99-011-0311A	4	4	4	
Clinical isolates New Zealand (ESR) [9]				
NZSPN00/9	4	4	4	
NZSPN00/42	18C	18C/18B	18C	

NZSPN00/59	5	5-q		
NZSPN00/87	13	13/20		
NZSPN00/88	6B	6B-g	6B	
NZSPN00/91	8	8	8	
NZSPN00/319	18B	18B/18C	18C	-; AY163212, AY163228
NZSPN00/366	7F	7F		
NZSPN00/426	3	3	3	
NZSPN00/454	23F	23F-23A= 23A-23F	23F	AF532679
NZSPN00/470	9V	9V	9V	
NZSPN00/480	6A	6A-ca	6B	
NZSPN00/484	23F	23F-g= 10A-23F	23F	
NZSPN00/499	19F	19F	19F	
NZSPN01/162	2	2-q	2	
NZSPN01/243	33F	33F-q	33F/37	-; AY163203, AY163219
NZSPN01/393	35F	35F		
NZSPN01/468	11A	11A-q		
NZSPN01/481	16F	16F		
NZSPN01/484	23F	23F-g= 10A-23F	23F	
NZSPN01/490	22F	22F		
NZSPN01/493	9N	9N		
NZSPN01/509	23A	23A-ca	23F-X; 23F-Y-NEG	
NZSPN01/510	12F	12F		
NZSPN01/520	9V	9V	9V	
NZSPN01/531	8	8	8	
NZSPN01/534	3	3	3	
NZSPN01/538	38	38/25F		
NZSPN01/543	10A	10A-q		
NZSPN01/546	4	4	4	
NZSPN01/547	20	20/13		
NZSPN01/548	7F	7F		
NZSPN01/549	1	1	1	
NZSPN01/553	17F	17F-c		
NZSPN01/554	19F	19F	19F	
NZSPN01/555	18C	18C/18B	18C	
NZSPN01/557	19A	19A	19A	
NZSPN01/559	6A	6A-c	6B	
NZSPN01/560	14	14-g	14	
NZSPN01/561	6B	6B-q	6B	
NZSPN00/12	17F	17F-c		
NZSPN00/50	Nonserotypeable	Nonserotypeable-riz		AF532714
NZSPN00/59	5	5-q		
NZSPN00/75	Nonserotypeable	No-amplicon		
NZSPN00/180	9V+14	9V	9V+14	
NZSPN00/221	38	38/25F		
NZSPN00/225	13	13/20		
NZSPN00/242	35F	35F		
NZSPN00/353	18A	18A	18C	AF532659; AY163209, AY163225

NZSPN00/410	33F	33F-q	33F/37	AF532688; AY163202, AY163218
NZSPN01/93	16F	16F		
NZSPN01/122	10A	10A-q		
NZSPN01/146	38	38/25F		
NZSPN01/166	16F	16F		AF532654
NZSPN01/204	35B	35B		
NZSPN01/209	22A	22A		
NZSPN01/240	12F	12F		
NZSPN01/254	35F	35F		
NZSPN01/262	8	8	8	
NZSPN01/276	6A	6A-6B-q =6B-q	6B	-; AY163177, AY163187
NZSPN01/278	18B	18B/18C	18C	-; AY163213, AY163229
NZSPN01/291	6B	6B-q	6B	
NZSPN01/303	Nonserotypeable	No-amplicon		
NZSPN01/313	18C	18C/18B	18C	
NZSPN01/329	6A	6A-6B-g =6B-g	6B	AF532701; AY163175, AY163185
NZSPN01/335	19A	19A	19A	
NZSPN01/344	18C	18C/18B	18C	
NZSPN01/361	9N	9N		
NZSPN01/363	18C	18C/18B	18C	
NZSPN01/366	6A	6A-ca	6B	
NZSPN01/369	18C	18C/18B	18C	
NZSPN01/374	35B	35B		
NZSPN01/387	22F	22F		
NZSPN01/388	12F	12F		
NZSPN01/389	20	20/13		
NZSPN01/403	20	20/13		
NZSPN01/411	11A	11A-nz		AF532638
NZSPN01/418	8	8	8	
NZSPN01/428	3	3	3	AF532683
NZSPN01/431	1	1	1	
NZSPN01/437	1	1	1	
NZSPN01/438	22F	22F		
NZSPN01/448	11A	11A-q		
NZSPN01/455	19A	19A	19A	
NZSPN01/463	10A	10A-q		
NZSPN01/465	22F	22F		
NZSPN01/477	10A	10A-23F =23F-g	23F-NEG	
NZSPN01/478	20	20/13		
NZSPN01/483	8	8	8	
NZSPN01/485	12F	12F		
NZSPN01/489	3	3	3	
NZSPN01/497	9N	9N		
NZSPN01/505	19A	19A	19A	
NZSPN01/512	7F	7F		
NZSPN01/515	3	3	3	
NZSPN01/516	1	1	1	
NZSPN01/529	1	1	1	
NZSPN01/532	4	4	4	
NZSPN01/535	7F	7F		

NZSPN01/539	19F	19F	19F
NZSPN01/545	18C	18C/18B	18C
NZSPN01/556	6B	6B-q	6B
NZSPN01/558	14	14-g	14

Notes.

1. CS of selected *S. pneumoniae* isolates from reference panels 1 and 3 was repeated by Gail Stewart and Robert Gange at Department of Microbiology, Children's Hospital at Westmead, New South Wales, Australia.
2. MCT was performed and GenBank accession numbers generated by Fanrong Kong at Centre for Infectious Diseases and Microbiology (CIDM), Institute of Clinical Pathology and Medical Research (ICPMR), Westmead Hospital, Westmead, New South Wales, Australia. See text for molecular capsular subtype (mctsp) nomenclature.
3. Provided by Denise Murphy, Pneumococcal Reference Laboratory, Public Health Microbiology, Queensland Health Scientific Services, Queensland, Australia.
4. Provided by Associate Professor Geoff Hogg and Jenny Davis, Microbiological Diagnostic Unit (MDU), Public Health Laboratory, Department of Microbiology and Immunology, University of Melbourne, Victoria, Australia.
5. Provided by Dr. Louise P. Jette, Institut National de Sante Publique du Quebec-Laboratoire de Sante Publique du Quebec, Sainte-Anne-de-Bellevue, Quebec H9X 3R5, Canada.
6. Provided by Dr. Michael Watson, Department of Microbiology, Children's Hospital at Westmead, New South Wales, Australia.
7. Selected 21 *S. pneumoniae* clinical isolates, of which CS results were known, from the CIDM diagnostic laboratory.

8. 152 Australian *S. pneumoniae* clinical isolates, of which CS results were known, from the CIDM diagnostic laboratory.

9. 103 New Zealand *S. pneumoniae* clinical isolates Provided by Dr. Diana Martin, from Streptococcus Reference Laboratory, at Institute of Environmental Science and Research (ESR), Wellington, New Zealand.

<u>Clinical isolates</u>

179 consecutive *S. pneumoniae* clinical isolates from normally sterile sites, collected during the period January 1999 to June 2001, by the CIDM diagnostic laboratory, were studied; 21 were randomly selected to make up reference panel 5 (see above). Dr Diana Martin, Institute of Environmental Science and Research (ESR), Wellington, New Zealand provided 103 clinical isolates from diagnostic laboratories throughout New Zealand. Clinical isolates were initially tested using the MCT method, without knowledge of their CS results (single-blind study). Isolates were retrieved from storage by subculture on blood agar plates (Columbia II agar base supplemented with 5% horse blood) and incubated overnight at 37°C CO_2 incubator.

<u>Conventional serotyping (CS)</u>

CS was performed by the Quellung reaction using rabbit polyclonal antisera from the Statens Serum Institute, Copenhagen, Denmark (Sorensen, 1993). Briefly, 2 μL of a suspension of isolate, in 10% formalin saline, and 1 μL of antisera, under a glass coverslip were examined for capsular swelling using a light microscope at 400x magnification. Clinical isolates from CIDM were serotyped at Department of Microbiology, Children's Hospital at Westmead, Sydney, Australia and those from New Zealand by the Streptococcus Reference Laboratory, at ESR, Wellington, New Zealand. Selected New Zealand clinical isolates for which only serogroup results were available and selected isolates from reference panels 1 and 3 were re-tested at Children's Hospital at Westmead.

<u>Molecular capsular typing (MCT) - development of method</u>

Oligonucleotide primers

The oligonucleotide primers used in this study, their target sites and melting temperatures are shown in Table 2 and the primer pair specificities and expected amplicon lengths in Table 3. Primers were designed with high melting temperatures to be used in rapid cycle PCR (Kong et al., 2000).

Table 2. Oligonucleotide primers used in this study.

Primer	Target gene	Tm °C[1]	GenBank accession numbers	Sequence[2-4]
*P1[5]	*psaA*	72.9	U53509	**203**TAC ATT ACT CGT TCT CTT TCT TTC TGC AAT CAT TCT TG**240**
*P2[5]	*psaA*	72.7	U53509	**1066**TAG TAG CTG TCG CCT TCT TTA CCT TGT TCT GC**1035**
*IIa[6]	*pneumolysin*	71.9	M17717	**457**AGA ATA ATC CCA CTC TTC TTG CGG TTG A**484**
*IIb[6]	*pneumolysin*	71.4	M17717	**680**CAT GCT GTG AGC CGT TAT TTT TTC ATA CTG**651**
cpsS1[7]	*cpsA (wzg)*	75.4	U09239	**1030**GGC ATT(/C) TAT GGA GTT GAT TCG(/A) TCC ATT(/C) CAC ACC(/T) TTA G**1066**
cpsS2[7]	*cpsA (wzg)*	71.9	U09239	**1057**CAC ACC(/T) TTA GAA AAT(/C) CTC TAT GGA GTG GAT ATC AAT TAC TAT G**1099**
cpsS3[7]	*cpsA (wzg)*	68.7	U09239	**1447**GAA AGT GGG(/A/T) GGG(/A/T) A(/G)A(/C)T(/G) TAT(/C) AAA GTA(/G) AAT TCT(/G) CAA GAT(/C) TTA(/G) AAA(/G) G**1489**
cpsA1[7]	*cpsA (wzg)*	71.5	U09239	**1549**CCA TCA C(/T)AT AGA GGT TAC(/A) TG(/A)T CTG GCA TT(/C)G C**1519**
cpsA2[7]	*cpsB (wzh)*	67.0	U09239	**1949**T(/G)CA TG(/A)C TA(/G)A AC(/T)T CT(/A)A TC(/T)A AG(/A)G CAT AAC GAC TAT C(/T)**1916**
cpsA3[7]	*cpsB (wzh)*	75.6	U09239	**2030**GC(/T)T CAA TG(/A)T GG(/A)G CAA TG(/T)A CTG GA(/C)G TA(/G)A TTC CCA(/G) ACA TC**1993**
1YS	*cap1H (wzy)*	72.1	Z83335	**10289**GTA GGT GTA GTT TTT TCA GGG ACT TTA ATT TTA TGC AGT G**10328**
1YA	*cap1H (wzy)*	70.4	Z83335	**10584** TCG CTT AAC ACA ATG GCT TTA GAA GGT AGA G**10554**
2YS	*cps2H (wzy)*	70.5	AF026471	**9711**GTT ATT TTA TTT TTT TTG TCG GCA TTG TAT TCT TTA TAT CG**9751**

2YA	*cps2H (wzy)*	71.3	AF026471	**10058**CAA ATT CAT CGT TTG TAT CCA TTT AAC TGC ATC**10026**
4YS	*wzy*	70.2	AF316639	**9601**CTT ATA TCT AAT TAT GTT CCG TCT ATA TTT ATA TGG GTT TGC TTT C**9646**
4YA	*wzy*	71.1	AF316639	**9948**TTT CTC TTC ATT TTC CTG ATA ATT TTG TAC TTC TGA ATG**9910**
6A6BYS0[7]	*wzy*	62.6	AY078347 & AF316640	**8196/9186**ATG CTT TTA AAT TTC TTA TTC ATA TCT ATT TTT C**8229/9219**
6A6BYS	*wzy*	72.0	AY078347 & AF316640	**8264/9254**G(/A)GA TTT T(/G)TT TCA ACC T(/C)GC AGT AAT TTT AAC AA(/C)T C(/T)G(/A)**8298/9288**
6A6BYA	*wzy*	71.4	AY078347 & AF316640	**8578/9568**CCT GAA AAC AA(/G)T ACT(/C) ACT TTC TGA ATT TCA C(/T)GG A(/G)TA TAA AG**8538/9528**
6A6BYA1[7]	*wzy*	72.4	AY078347 & AF316640	**8944/9934**GTA AAC AGA GAG CGA GTG ATC ATT TTA AAA CTT TTG G**8808/9898**
8YS	*wzy*	70.5	AF316641	**10810**GTT TTA TTG ACT TTA AAG ATG TTA GTT TCT TCG ATT CCA G**10849**
8YA	*wzy*	70.5	AF316641	**11086**TTT TTA TTA CTC TTC TTA AAT CAT AAT GAA TCG TAC CAA TCA AC**11043**
9VYS	*cps9vI (wzy)*	73.5	AF402095	**8535**GGA TCA ATG GCA ACT ATA TTT ACC CTA CTC TCC ACA G**8571**
9VYA	*cps9vI(wzy)*	76.3	AF402095	**8872**GAG TCG AAA CCA ACC GGA AAA AGC AAT TGA G**8842**
14YS	*cps14H (wzy)*	71.5	X85787	**7361**CCT TTG GTT TAT TAT CCT ACT TCC AAA ACA GTT TAT GC**7398**
14YA	*cps14H (wzy)*	71.4	X85787	**7670**CAT ATA TCT CTT TAT CCT GTC AAT ATT GAT TGG CAT TTT C**7631**
18CYS0[7]	*wzx*	71.3	AF316642	**11856**GAA ATT ATA GTC GGA GCT TTC ATT TAT ATT AGT TTA CTG GTT CTG**11900**
18CYS	*wzy*	71.5	AF316642	**12190**GAT ATT AGC TAT ACC AAC AAT TGT TCT TTT CCT GTA CTC AGT C**12232**
18CYA	*wzy*	72.5	AF316642	**12491**GCA TTT CTA GTA CCG AAC CAT TGA AAC TAT CAT CTG**12456**

18CYA1[7]	*wzy*	73.3	AF316642	**12536**CAG AAT AAA GAG AGC TGT AAT AGG TGC AAC TTC ATG C**12490**
19FYS	*cps19fI (wzy)*	70.6	U09239	**7673**CTG TAA TGT TTC TAA TTA GTT CAG TAT TTG CAC TGG TTA ATT C**7715**
19FYA	*cps19fI (wzy)*	72.0	U09239	**7958**CCC GTA TAT CCA TTA CTA AGA ACA AGG TTG TAT ATT TCC TTC**7917**
19AYS	*cps19aI (wzy)*	71.2	AF094575	**9245**GTT TCT CAT TAG TTC TGT ATT TGC CCT TAT TAA TGT GC**9282**
19AYA	*cps19aI (wzy)*	72.2	AF094575	**9514**CCA TGG CTA AGT GCA AGA TTA TGA ATC TCT CTC**9482**
19BYS	*cps19bI (wzy)*	71.6	AF004325	**3519**GTT TCT TAT GTT TAC CCT CAG CTT ATA TTG GCA CAG**3554**
19BYA	*cps19bI (wzy)*	71.5	AF004325	**3946**GAT ACC ACA AAT CTC CGA ATT CTC TTA AAA TAG ATG G**3910**
23FYS	*cps23fG (wzy)*	71.6	AF057294	**8567**TTA AGT AGT TCA CAA GTG ATA GTG AAC TTG GGA TTG TC**8604**
23FYA	*cps23fG (wzy)*	70.7	AF057294	**8846**CAC TGA GAT TAT TTA TTA GCT TTA TCG GTA AGG TGG ATA AG**8806**
33F37YS0[7]	*cap33fJ*	76.0	AJ006986	**11191**CCA ATG AAA AGG AAA GTT CAA TGT GTT TTG TTT CTG C**11227**
33F37YS	*cap33fK & cap37K (wzy)*	70.7	AJ006986 & AJ131984	**11341/11708**ATT ACT TGT AAT ACT ATG TAT TCA ACT AGT CA(/C)A GGA TTT GAT GG**11384/11751**
33F37YA	*cap33fK & cap37K (wzy)*	71.7	AJ006986 & AJ131984	**11650/12017**GAACAAATTTCCGTATCAGATTTGCGATT TC**11620/11987**
33F37YA1[7]	*cap33fK (wzy)*	72.2	AJ006986	**11858**GGT GCT TCA GCA AAA ATC CCC GTA TTT CTT ATC AG**11824**
1XS	*cap1I (wzx)*	72.6	Z83335	**12017**TAG CTG ATG TTC CGA TAA ATT ATG GTG GGG TAA TAA TAG**12055**
1XA	*cap1I (wzx)*	70.6	Z83335	**12442**CTG CGA CAC TGT ATA TAC CTA CAT TAT AAC TAC TAG ACA TTT GC**12399**
2XS	*cps2J (wzx)*	71.8	AF026471	**12167**GCA ACT TTG GTT CTA AAA TTT TAG TCT TTT TAA TGG TTC C**12206**

2XA	*cps2J (wzx)*	72.1	AF026471	12595TGT TAA ACC CCA ATA TAG AAA TTG TAT TGA GAA TAG CAG C12556
4XS	*wzx*	73.2	AF316639	12119CG TTA ATA GCT TAT GTT CAA CTG GTG ATT GAT TTT GG12155
4XA	*wzx*	72.0	AF316639	12442TGA TAG TTT TAG AAA TAA TAT AAG GAA TTG CAA CTG CAT GC12402
6A6BXS0[7]	*cpsI-wzx* spacer	72.7	AY078347& AF246898	9581/4550GGT AGG TAT TTT AAT TGG AGG AAG AGA GTC TTG AAT GG9618/4587
6A6BXS	*wzx*	72.5	AY078347 & AF316640	9695/10685TTC ATG TC(/T)T(/C) TTT TG(/A)T CTA ATC TGA TTA CAA TTG(/C) TC(/T)A CAT CG(/A)9735/10725
6A6BXA	*wzx*	74.1	AY078347 & AF316640	9999/10989T(/C)GC ATT TG(/T)G ATC TGT CAC AA(/G)T CAA TAA GTT AAA ACC9964/10954
6A6BXA1[7]	*wzx*	72.5	AY078347& AF246898	10682/5651ATC TTC CCT TCA TAA ATT GAC ATA GGA AAA ATA AGA GCC10644/ 5613
8XS	*wzx*	71.8	AF316641	8602CAA TTC TAA CTA TGT CCA GTT TTA TTT TTC CAC TCA TCA G8641
8XA	*wzx*	74.2	AF316641	8926GAC GTG ATA ATA ATA AGC TGC CAT TCC TGT CTA AAA CG8889
9VXS	*cps9vK (wzx)*	74.5	AF402095	10543CGG CGG TAT TAA GTA GAA TAT TAA CAC CTG AAG AGT ATG GC10583
9VXA	*cps9vK (wzx)*	73.6	AF402095	10910GGC AAT CAG ACT CAA TAA GTT CAT CCG TTT AAA GTT C10874
14XS	*cps14L (wzx)*	72.1	X85787	11463GGT ATT GCC TTT CCT TTG ATA ACT TCT CCT TAT TTA TCA C11502
14XA	*cps14L (wzx)*	71.6	X85787	11751TGA ACT TGT AAC TCG ACA CCC AAA AAT ATA AAT AAA TGA G11712
18CXS0[7]	*wciW*	75.0	AF316642	10403CAA AGG AAC GTT ATC AGC AAT TGT GTC AAA TTT CAG10438
18CXS	*wzx*	72.5	AF316642	10715GAA TCG GAC AAT AGC ACA GGT ACG AAC AAG10744
18CXA	*wzx*	75.2	AF316642	11082GCC ATG TAA TCA ACT GAC CAA GCA GGG TAC TC11051

18CXA1[7]	*wzx*	72.2	AF316642	**11123**AAG ATT AGG GCG CAC AAA GTT TAC TTG TTT TAG C**11090**
19FXS	*cps19fJ (wzx)*	71.3	U09239	**8975**GTT ATT TCT TCA AAT CTG CTC ATA GTT TTA ACC TCA TCA C**9014**
19FXA	*cps19fJ (wzx)*	73.5	U09239	**9279**TAT CTT GCG TTT TCA TCC CTT ACA GTT ATT AGG TTC AAA G**9240**
19AXS	*cps19aJ (wzx)*	74.7	AF094575	**10547**TTC TTC AAA TCT TTT GAC AGT CTT GAC CTC TTC CTT G**10583**
19AXA	*cps19aJ (wzx)*	72.3	AF094575	**10846**TAT CGT GCA TTC GAA TCT GTT ACA GCT AAT ACA TTT AAA C**10807**
19B/19CXS	*cps19bJ (wzx)*	74.3	AF004325 & AF105116	**7778/373**GTC CTG ACG CTA TCA AAT ATC ATT TTC CCA TTA ATC AC**7815/410**
19B/19CXA	*cps19bJ (wzx)*	73.2	AF004325 & AF105116	**8104/699**CCC ACA TGT GAT CAA TAG GAG TGA AAA TTC TCT ATT C**8068/663**
23FXS0[7]	*cps23FI*	73.4	AF057294	**11714**CCT TTG GCT AAT TTC TTG GAC GAT AAT GAA TTT GTA TAT G**11753**
23FXS	*cps23fJ (wzx)*	72.3	AF057294	**11961**GCT TTG GCT AAC TTT TCA TCA AAG ATT TTA ATT TTT TTG TTA G**12003**
23FXA	*cps23fJ (wzx)*	73.3	AF057294	**12361**CCA GAG ATA GCT GTA ACA CCA ATT TTA TCA ATT CCC TTA G**12322**
23FXA1[7]	*cps23fJ (wzx)*	72.5	AF057294	**12457**CCA CAA ACA TTA GCA ATA AAG AAA CCT AAC AAT CCC**12422**
33F37XS0[7]	*cap33fK (wzy)*	76.7	AJ006986	**12271**GTT GTT TTA GCT CAA GGA GGG ATA ATG TTG GCT TCG**12306**
33F37XS	*cap33fI & cap37L (wzx)*	72.2	AJ006986 & AJ131984	**12591/12958**GAT CAT ACT CCC TAT CAT TAC GAC TCC CTA TGT AAC G**12627/12994**
33F37XA	*cap33fI & cap37L (wzx)*	72.1	AJ006986 & AJ131984	**12918/13285**CCA AGA AAT ATC CAA ACC TTT TGA CAC TAA ACT TAA TCC**12880/13247**
33F37XA1[7]	*cap33fL (wzx)*	73.3	AJ006986	**13016**GCT GAT TTT ACA AAT AGG AAA ATA GAG ATT GCA CCA AC**12979**
3S1	*orf2 (wze)- cap3A* spacer	72.6	Z47210	**5793**GCA CAA AAA AAA GTT TGA TAT TCC CCT TGA CAA TAG**5828**

3A1	*cap3A*	73.3	Z47210	6113GCA GGA TCT AAG GAG GCT TCA AGA TTC AAC TC6082
3S2	*cap3A*	72.4	Z47210	6933CGA ACC TAC TAT TGA GTG TGA TAC TTT TAT GGG ATA CAG AG6973
3A2	*cap3B*	75.7	Z47210	7229CTG ACA GCA TGA AAA TAT ATA ACC GCC CAA CGA ATA AG7192

Notes.

1. Primer Tm values provided by the primer synthesiser (Sigma-Aldrich).
2. Numbers represent the numbered base positions at which primer sequences start and finish (starting at point "1" of the corresponding gene GenBank sequence).
3. Underlined sequences show bases added to modify previously published primers.
4. Letters in parentheses indicate alternative nucleotides in different serotypes.
5. Morrison, et al. 2000 (24).
6. Salo, et al. 1995 (27).
7. For sequencing use only.

* Primers have been previously published. All others primers designed specifically for this study

Table 3. Specificity and expected lengths of amplicons of primer pairs used in this study.

Primer pairs[1]	Specificity	Length of amplicons (base pairs)
P1/P2	*S. pneumoniae*	864
IIa/Iib	*S. pneumoniae*	224
cpsS1/cpsA3[2]	*S. pneumoniae*	1001
cpsS1/cpsA1[2]	*S. pneumoniae*	520
cpsS3/cpsA2[2]	*S. pneumoniae*	503
1YS/1YA	serotype 1	296
2YS/2YA	serotype 2	348
4YS/4YA	serotype 4	348
6A6BYS/6A6BYA	serogroup 6	315
6A6BYS0/6A6BYA1[2]	serogroup 6	747
8YS/8YA	serotype 8	277
9VYS/9VYA	serotype 9V	338
14YS/14YA	serotype 14	310
18CYS/18CYA	serogroup 18	302
18CYS0/18CYA1[2]	serogroup 18	671
19FYS/19FYA	serotype 19F	286
19AYS/19AYA	serotype 19A	270
19BYS/19BYA	serotype 19B	428
23FYS/23FYA	serotype 23F	280
33F37YS/33F37YA	serotypes 33F/33A/37	310
33F37YS0/33F37YA1[2]	serotypes 33F/33A/37	668
1XS/1XA	serotype 1	426

2XS/2XA	serotype 2	429
4XS/4XA	serotype 4	324
6A6BXS/6A6BXA	serogroup 6	305
6A6BXS0/6A6BXA1[2]	serogroup 6	1102
8XS/8XA	serotype 8	325
9VXS/9VXA	serotype 9V	368
14XS/14XA	serotype 14	289
18CXS/18CXA	serogroup18	368
18CXS0/18CXA1[2]	serogroup 18	721
19FXS/19FXA	serotype 19F	305
19AXS/19AXA	serotype 19A	300
19BXS/19BXA	serotype 19B	327
23FXS/23FXA	serotypes 23F/23A	401
23FXS0/23FXA1[2]	serotypes 23F/23A	744
33F37XS/33F37XA	serogroups 33/37	328
33F37XS0/33F37XA1[2]	serotypes 33F/33A/37	746
3S1/3A1	serotype 3	321
3S2/3A2	serotype 3	297

Notes.

1. See Table 2 for primer sequences.
2. For sequencing use only.

Four previously published *S. pneumoniae*-specific primers, targeting psaA (P1, P2) (Morrison et al., 2000) and pneumolysin (IIa, IIb) (Salo et al., 1995) were modified to give high melting temperatures and used to confirm that isolates were *S. pneumoniae*. Primers were designed to amplify and sequence portion of the *cpsA-cpsB* gene region and to amplify serotype/serogroup-specific sequences in the *wzy* and *wzx* genes of 16 *S. pneumoniae* serotypes for which cps gene cluster sequences were available. In order to further explore the sequence heterogeneity, part of the *wzx* and *wzy* genes of isolates belonging to serogroups 6, 18, 23 and 33/37 were also sequenced. For serotype 3, which does not contain *wzy* and *wzx* genes, serotype-specific PCR targeted the orf2 (wze)-cap3A-cap3B region (Arrecubieta et al., 1996).

DNA preparation, PCR and sequencing

DNA extraction, PCR and sequencing were performed as previously described (Kong et al., 2002).

Sequence comparison, multiple sequence alignments, and phylogenetic analysis

Sequences were compared using Bestfit in Comparison program group. Multiple sequence alignments were performed with Pileup and Pretty in Multiple Sequence Analysis program group. Phylogenetic relationships were studied using Ednadist and Ekitsch in Evolutionary Analysis program group. All programs are provided in WebANGIS, ANGIS (Australian National Genomic Information Service), 3rd version.

Nucleotide sequence accession numbers

The new partial sequence data for *cpsA-cpsB*, *wzy* (polymerase) and *wzx* (flippase) genes for selected reference and clinical isolates reported in this paper have appeared in the GenBank Nucleotide Sequence Databases, with accession numbers AF532632-AF532715, and AF163171-AF163232, respectively (Table 1).

Previously reported sequence data used in this paper, in addition to those listed in Table 2, have appeared in GenBank Nucleotide Sequence Databases with the following accession numbers: U15171, U66846 and U66845 (*cps* gene cluster for serotype 3); NC_003028 (serotype 4 genome); AJ239004 (*cps* gene cluster for serotype 8); AF030367-AF030372 (*cps* gene cluster for serotype 19F); AF105113 (partial *cps* gene cluster for serotype 19A); AF105114 and AF106137 (partial *cps* gene clusters for serotype 19B); AF105115 (partial *cps* gene clusters for serotype 19C); AF030373 and AF030374 (*cps* gene clusters for serotype 23F).

RESULTS

Both pairs of *S. pneumoniae* species-specific primers (targeting psaA and pneumolysin genes) produced amplicons of the expected size from all reference and clinical isolates except six of 179 CIDM isolates, which, on retesting, were optochin resistant and therefore excluded from further study as they were not *S. pneumoniae*.

The sequencing primers, cpsS1/*cpsA*3, formed amplicons from all but 13 reference and clinical isolates. Of these 13 isolates, 10 (eight belonging to serotypes 38/25F and two that were nonserotypable) formed amplicons with primer pairs cpsS1/*cpsA*1 and cpsS3/*cpsA*2. Three nonserotypable isolates did not form amplicons using any of the primer pairs targeting the *cpsA-cpsB* region, although they had been confirmed to be *S. pneumoniae* using both species-specific PCR.

<u>Sequence heterogeneity in the region between the 3'-end of *cpsA and* the 5'-end of *cpsB*</u>

The present inventors sequenced and analyzed 800 bp fragments of the region between the 3'-end of *cpsA* (starting at base pair 951) and the 5'-end of *cpsB* (see Figure 2). Representative sequences were deposited into GenBank (see Table 1 for accession numbers). There were 424 sites that were identical for all 51 serotypes represented among the isolates examined, leaving 376 (47%) heterogeneity sites.

<u>Intra- and inter-serotype/subtype heterogeneity</u>

Only single isolates were available for 11 serotypes and the mixed serotype 9V/14 (see below). Among 40 serotypes, for which multiple isolates were available, 14 were divided into two or more subtypes, on the basis of major and/or stable intra-serotype heterogeneity. Molecular capsular subtypes (mcst) were named according to their conventional serotype (cs) and, generally, the source of the isolate in which the sequence difference was first identified [-g = Genbank sequence; -c (CIDM); -q (Queensland); - ca (Canada); -nz (New Zealand)]. When sequences characteristic of two serotypes were present in the *cpsA-cpsB* region subtype names included both, with the cs first (e.g mcst 23F-23A when cs was 23F; mcst 23A-23F when cs was 23A). Seventeen serotypes had no intra-serotype heterogeneity and in nine there were minor and/or less stable variations between isolates and/or between sequences disclosed herein with corresponding sequences in GenBank (Table 4, Figure 2).

There were 368 heterogeneity sites that allowed differentiation between molecular capsular types (mct) and subtypes (mcst), including both specific and shared sites (Table 4, Figure 2).

Table 4. Intra-molecular capsular type/subtype (mct/mcst) and inter-mct/mcst heterogeneity sites in the region between the 3'-end of *cpsA* and the 5'-end of *cpsB* of 51 *S. pneumoniae* serotypes.

Mct/mcst (n=)[a]	Intra-mct/mcst[b] Heterogeneity Site – base	Identity between mcst (%)	Mct/mcst[b]-specific heterogeneity site – base	Selected heterogeneity sites shared with other mct/mcst[b]- base
1 (9+g)	133 - T^{8}/A^{9}		289 - A, 452 – A	122 - T, 152 - A, 495 - A, 600 - A
2-g (g)	-		705, 706 – CG	287 - G, 507 - G, 534 - A
2-q (3)	Nil	95.9%	239 - C, 293 - T, 386 - A, 404 – G	232 - G, 286 - C, 600 - A
3 (17+g)	262 - C^{g+16}/T^{1}, 292 - G^{16}/A^{g+1}, 293 – A^{16}/G^{g+1}, 539 – C^{16}/T^{g+1}, 545 - C^{g+16}/A^{1}		485 - A, 487 – A	27 - A, 90 - A, 231 - A, 590 - T, 686 - T
4 (36)	Nil		179 - C	231, 232 - TG, 611 - T, 743 - T
5-q (4)	Nil			428 - T, 599 - A
5-c (1)	-	94.0%		122 - T, 152 - A, 247 - C, 605 - T

6A-g (g)	463-5 – AGC^{12}/GCA^{8}, 534 - A^{8}/G^{12}, 542 - C^{8}/T^{12}, 545 - A^{8}/C^{12}			62 - A, 209 - A, 534 - A, 542 - C
6A-ca (7)	55 - A^{5}/G^{2}, 331 - A^{2}/G^{5}, 434 - A^{5}/G^{2}	6A-ca : 6A-g=99.1%		62 - A, 209 - A
6A-c (2)	Nil	6A-c : 6A-ca= 99.5%		62 - A, 209 - A, 337 - G
6A-6B-g (2)	(see 6B-g) 772 - A^{8+1}/G^{1}			(see 6B-g)
6A-6B-q (1)	(see 6B-q)			(see 6B-q)
6B-g (4+g)	31 - A^{1}/G^{8+3}			209 - A, 337 - G, 341 - G,
6B-q (9)	383 - A^{8}/G^{1}	6B-q : 6B-g=84.7%	749 - G	52 - G, 58 - C, 68 - G, 82 - C, 85 - T, 94 - T, 104 - T, 116 - G, 160 - T, 209 - C, 286 - C, 343 - G, 375 - G, 478 - C, 490 - C, 521 - T, 563 - T, 704 - C, 776 - C
6B-c (1)	-	6B-c : 6B-g=92.1%		193 - T, 209 - C
7F (15)	Nil		66 - C, 445 - C	722 - C, 731 - A
7C (3)	Nil			49 - C, 731 - A

8 (12)	Nil		340 - T, 670 - G	425 - A
9N (9)	Nil		81 - T, 378 - A	352 - G, 409 - T, 590 - T, 722 - A
9V (17)	Nil		245 - G	428 - C, 704 - C, 750 - T, 776 - C
10F (2)	309 - G^1/A^1, 335 - G^1/A^1			704 - C, 750 - T, 776 - C
10A-q (5)	Nil		222 - T, 663 - T	232 - G
10A-23F (6)	(see 23F-g)	91.2%		(see 23F-g)
11A-q (7)	Nil			122 - T, 232 - G, 478 - C, 490 - C, 521 - T, 704 - C
11A-nz (1)	-	94.0%	316 - T	597 - A
11B (1)	-		269 - A, 490 - G, 776 - T	10 - G, 52 - G, 58 - C, 68 - G, 82 - C, 85 - T, 94 - T, 104 - T, 116 - G, 148 - T, 160 - T, 231, 232 - TG, 247 - C, 250 - A, 286 - C, 292 - C, 343 - G, 375 - G, 425 - A, 521 - T, 563 - T, 704 - C
12F (9)	268 - A^1/C^8, 572 - C^1/T^8, 781 - G^1/T^8		274 - C	287 - G, 497 - G, 577 - T, 722 - C
13 (6)/20 (8)	Nil; Nil			590 - T, 686 - T, 722 - A

14-g (32+g)	249 - T^{23}/C^{8+9}, 250 - G^{32}/T^{8}, 320 - G^{32}/A^{8}			577 - T
14-c (1)	-	98.1%	613 - G	16 - C, 49 - C, 54 - T, 62 - T, 406 - G, 577 - T
15A-ca1 (1)	-		473 - G	49 - C, 337 - G, 507 - G
15A-ca2 (1)	-	95.1%	406 - A, 473 - G	337 - G, 507 - G
15B-q (5)	Nil			232 - G
15B-c (1)	-	15B-c : 15B-q=97.4%	235 - T, 351 - G	49 - C, 247 - C, 352 - G, 428 - T, 542 - C
15B-22F (2)	(see 22F)	15B-22F : 15B-q=95.2%		(see 22F)
15C-q (1)	as for 15B-q plus 104 - T^{C}/C^{B}			232 - G
15C-CA (1)	as for 15B-q plus 232 - A^{C}/G^{B}, 757 - T^{C}/C^{B}	99.6%		pattern
16F (6)	149 - C^{5}/T^{1}, 232 - A^{5}/G^{1}			122 - T, 232 - G, 352 - G, 548 - A
17F-c (3)	Nil			199 - A, 247 - C, 600 - C
17F-35B (2)	(see 35B)	99.8%	728 - C	(see 35B)
17A (1)	-		122 - A	85 - T, 554 - G, 567 - A

18F (1)	-	65 - A, 161 - T, 469 - C, 684 - A	722 - C, 786 - C
18A (2)	63 - T^{1}/A^{1}	99 - C, 202 - G, 232 - C, 239 - G, 322 - C, 334 - C,	122 - T, 307 - G, 563 - T, 686 - T
18B (4)/18C (14)	Nil; Nil	138 - G, 459 - C, 750 - A	478 - C
19F (20+gx7)	164 - C^{gx7+17}/T^{3}, 169 - C^{gx6+11}/T^{g+9}, 387 - A^{gx6+20}/T^{g}, 414 - G^{gx5+20}/T^{gx2}, 479- G^{gx7+16}/A^{4}		169 - T, 337 - G
19A (11+g)	70 - T^{8}/C^{11}, 479 - A^{8}/G^{g+3}	202 - C	49 - C, 54 - T, 62 - T, 94 - A, 103 - C, 104 - T, 160 - T, 198 - C, 232 - G, 286 - C, 343 - G, 352 - G, 375 - T, 425 - A, 490 - C, 750 - T
21 (1)	-		428 - C, 548 - A, 629 - T, 717 - A
22F (13)	Nil		428 - T, 567 - G, 599 - A, 731 - A
22A (4)	Nil		428 - T, 567 - A, 599 - A, 731 - A
23F-g (17+gx3)	Nil		193 - T

23F-c (1)	-	23F-c : 23F-g=91.2%	88 - G	249 - A, 337 - G
23F-23A (1)	-	23F-23A : 23F-g=98.7%		495 - A
23A-ca (2)	Nil			247 - C, 495 - A
23A-23F (1)	(as for 23F-23A)	96.6%		(as for 23F-23A)
23B (1)	-		734 - C, 763 - G	49 - C, 55 - T, 58 - C, 62 - T, 103 - C, 104 - T, 160 - T, 198 - C, 223 - G, 232 - G, 249 - T, 286 - C, 292 - C, 343 - G, 375 - G, 376 - G, 425 - A, 490 - C, 521 - T, 563 - T, 704 - C
25F (1)/38 (7)	-; Nil		Numerous sites	Numerous sites
29 (1)	-		310 - A	335 - A
31 (2)/42 (1)	Nil; -			122 - T, 152 - A, 605 - T
33F-g (2+g)/33A (1)	534 - A^8/G^2 ; -			247 - C, 600 - A, 728 - T
33F-q (4)	313 - T^1/G^3	94.7%	313 - T	169 - T, 717 - A
33B (1)	-		578 - G	169 - T, 717 - A
34 (2)	Nil			85 - C, 122 - C, 554 - G, 567 - A

35F (6)	Nil		232 - G, 343 - G, 554 - G, 577 - T
35B (9)	Nil		199 - G, 247 - C, 600 - A, 728 - C
37 (1+g)	231 - A^g/C^1	54 - G	90 - A, 231 - A, 743 - T
41F (1)	-		287 - G, 507 - G

Notes.

a. Key to mcst: -g = Genbank sequence; -c (CIDMLS); -q (Queensland); - ca (Canada); -nz (New Zealand)

b. The superscript numbers = number of isolates studied; superscript g = base present in corresponding GenBank sequence

c. Site 567 heterogeneity (A/G) can correctly distinguish serotypes 22F (G) and 22A (A).

Phylogenetic tree based on region of the 3'-end of *cpsA*-the 5'-end of *cpsB* genes

Using these 800bp sequences, a phylogenetic tree was inferred for the 71 *S. pneumoniae* mct and mcst (Figure 65). *S. pneumoniae* can be divided into at least two classes, based on sequence analysis of the *cps A-D* region. Typical class I serotypes (e.g. 1, 18C, 19F), a typical class II serotype (e.g 33F, represented by 33F-g) and a nontypical class II serotype (19A) were each in different clusters of the tree (Jiang et al., 2001).

The phylogenetic tree provides evidence for, and suggests possible sources of, recombination between *cpsA-cpsB* genes of classes I and II. For example, subtype 23F-c clustered with 15A-ca2, but in a separate cluster from other 23F and 15A subtypes, suggesting that they may have arisen by recombination between 23F and 15A, respectively, and other serotypes. Different subtypes of some other mct were located in different clusters and appeared to be only distantly related to each other e.g. 33F-g and 33F-q, 2-g and 2-q, 17F-c and 17F-35B. Sharing of identical sequences between otherwise unrelated serotype pairs also provides evidence of recombination (see above).

Molecular capsular typing (MCT) based on *cpsA-cpsB* region sequences

The mct, assigned on the basis of *cpsA-cpsB* sequence, was the same as the cs for all isolates belonging to 36 of 51 serotypes (or 304 of 394 [77%] isolates), and for the majority of isolates (25 of 39) belonging to another five serotypes (Table 5). The remaining isolates in these serotypes shared sequences with other serotypes, namely 6A with 6B, 10A and 23A with 23F, 15B with 22F and 17F with 35B, presumably as a result of recombination. There were five serotype pairs, represented by 46 isolates, whose members had identical sequences: namely 20/13, 18C/18B, 38/25F, 31/42 and 33F-g/33A.

MCT based on PCR targeting *wzy* and *wzx* (*orf2 [wze]-cap3A-cap3B* for serotype 3)

There is significant sequence heterogeneity in *wzy* and *wzx* (data not shown), which made them suitable PCR targets for serogroup or serotype identification (Tables 2 and 3). With few exceptions, primer pairs targeting these genes formed amplicons only from the corresponding serotypes represented in the five reference panels. Exceptions were: PCR targeting serotype 6B also amplified 6A; PCR targeting 18C amplified all serotypes in serogroup 18; PCR targeting *wzx* (but not *wzy*) of serotype 23F, amplified three serotype 23A strains; PCR targeting *wzx* and *wzy* of serotypes 33/37 amplified a 33A isolate and that targeting *wzx* amplified a serotype 33B isolate.

The specificity of serotype 3-specific primers targeting the orf2 (wze)-cap3A-cap3B genes (Arrecubieta et al., 1996) was confirmed by production of an amplicon of the expected size from all 17 serotype 3 isolates. Thus, a serotype or serogroup was assigned by PCR to all 239 isolates belonging to serotypes/serogroups for which specific PCR was developed (Table 5).

Comparison of MCT based on *cpsA-cpsB* sequencing and PCR/sequencing targeting *wzx* and *wzy*

The results of PCR and *cpsA-cpsB* sequencing were consistent except that PCR could not distinguish between some members of serogroups 6, 18, 23 and 33/37 and further sequencing (of *wzx, wzy*) was required to identify individual mct/mcst (see below). The *cpsA-cpsB* sequences of six 10A isolates were identical to those of 23F, but the isolates were negative in the 23F-specific PCR targeting *wzx* and *wzy* (mcst 10A-23F).

Relationships within serogroups

Sequence analysis of the *cpsA-cpsB* region and *wzy* and *wzx* genes (data not shown) showed variable phylogenetic relationships between members of different serogroups.

Serogroup 6

Mct 6A and 6B were divided into five and three subtypes, respectively, based on different sequence patterns in the *cpsA-cpsB* region. Three 6A isolates had sequences in this region characteristic of serotype 6B (Table 4). Serotypes 6A and 6B could not be distinguished by PCR targeting *wzx* and *wzy*. Sequencing of these genes correctly identified all except one 6A isolates, but some 6A and 6B subtypes share identical or very similar sequences. The serotype of the discrepant isolate (serotype 6A, mcst 6B-q) was checked independently by two laboratories (Vakevainen et al., 2001).

Serogroup 18

Mct 18C and 18B had identical *cspA-cpsB* region sequences and were close to 18A and 18F in the class I cluster (Figure 65). PCR targeting both *wzx* and *wzy* genes amplified all four serotypes. Sequences of 18C and 18B were identical to each other, but different from those of serotypes 18A and 18F, which were also distinguishable from each other.

Table 5. Comparison of molecular capsular typing (MCT) and conventional serotyping (CS) results of 394 *S.pneumoniae* isolates.

CS	N=	mct-seq: a) *cpsA-cpsB* or b) *wzx, wzy* type(s) (n)[1]	mct-PCR *(wzy & wzx)*	Final mct	Comment
1	9	1	1	1	Correlate
2	3	2	2	2	"
3	17	3	3	3	"
4	36	4	4	4	"
5	5	5	NA	5	"
6A	12	a) 6A(9); 6B-g (2); 6B-q (1)	Serogroup 6	6A (11)	1 of 12 results
		b) 6A (11)[2]; 6B-q (1)		6B (1)	discrepant[2]
6B	15	6B	Serogroup 6	6B	Correlate
7C	3	7C	NA	7C	"
7F	15	7F	NA	7F	"
8	12	8	NA	8	"
9N	9	9N	NA	9N	"
9V	17	9V	9V	9V	"
9V/14	1	9V	9V/14	9V/14	See text

10A	11	10A (5); 23F-g (6)[3]	23F *wzy/wzx* PCR negative (6)[3]	10A (11)[3]	Correlate[3]
10F	2	10F	NA	10F	Correlate
11A	8	11A	NA	11A	"
11B	1	11B	NA	11B	"
12F	9	12F	NA	12F	"
13	6	13/20	NA	13/20	Consistent
14	33	14	14	14	Correlate
15A	2	15A	NA	15A	Correlate
15B	8	15B (6); 22F (2)	NA	15B (6); 22F (2)	2 of 8 results discrepant
15C	2	15C	NA	15C	Correlate
16F	6	16F	NA	16F	"
17A	1	17A	NA	17A	"
17F	5	17F (3); 35B (2)	NA	17F (3); 35 (2)	2 of 5 results discrepant
18A	2	18A	Serogroup 18	18A	Correlate
18B	4	18C/18B	"	18B/C	Consistent
18C	14	C/18B	"	18B/C	"

18F	1	18F	"	18F	Correlate
19A	11	19A	19A	19A	"
19F	20	19F	19F	19F	"
20	8	13/20	NA	20	Consistent
21	1	21	NA	21	Correlate
22A	4	22A	NA	22A	"
22F	13	22F	NA	22F	"
23A	3	a) 23A (2); 23F-g (1) b) 23A (3)[4]	23F *wzy* PCR negative/23F *wzx* PCR positive[4]	23A[4]	"[4]
23B	1	23B	NA	23B	"
23F	20	23F	23F	23F	"
25F	1	25F/38	NA	25F/38	Consistent
29	1	29	NA	29	Correlate
31	2	31/42	NA	31/42	Consistent
33A	1	33A/33F-g[5]	Serogroup 33/37[5]	33A/33F[5]	"[5]
33B	1	33B	Serogroup 33/37 PCR (*wzy*) negative[6]	33B	Correlate[6]

33F	6	33A/33F-g[5], 33F-q	Serogroup 33/37[5]	33A/33F[5]	Correlate[5]
34	2	34	NA	34	Correlate
35B	9	35B	NA	35B	"
35F	6	35F	NA	35F	"
37	1	37	Serogroup 33/37	37	"
38	7	25F/38	NA	25F/38	Consistent
41F	1	41F	NA	41F	Correlate
42	1	31/42	NA	31/42	Consistent
Nonserotypable	5	Non-typable[7]	NA[7]	Non-typable[7]	Correlate[7]
TOTAL	394				**Results:** Correlate = 343 Consistent = 46 Discrepant =5

Notes.

1. For mcst nomenclature, see Table 4 and text.

2. *cpsA-cpsB* sequence 3 discrepancies; 2 resolved by *wzx*, *wzy* gene sequences.

3. Six serotype 10A isolates shared *cpsA-cpsB* sequence with 23F-g, but 23F specific PCR (targeting both *wzy* and *wzx*) was negative;mcst10A-23F was identified by exclusion of 23F in our existing database. However, this relationship needs to be confirmed by examination of alarger collection isolates.
4. *cpsA-cpsB* sequence 1 discrepancy; resolved by *wzx* gene sequence; 23F *wzx* PCR positive/23F negative *wzy* PCR negative also support its identification by exclusion.
5. For one serotype 33A isolate, *cpsA-cpsB* and *wzx* and *wzy* sequences were identical with 33F-g but different from 33F-q; 33F/37 *wzx* and *wzy* PCR were both positive.
6. One serotype 33B strain identified by exclusion: 33F/37 *wzx* PCR positive/33/37 *wzy* PCR negative.
7. All isolates confirmed to be *S. pneumoniae*. These isolates may belong to rare serotypes not represented among our reference isolates.

Serogroup 23

Mct 23F, 23A (except mcst 23F-23A and 23A-23F) and 23B were separated into different clusters based on *cpsA-cpsB* sequence differences. Serotype 23A (including mcst 23A-23F) was identified on the basis of a positive result with 23F-specific primers targeting *wzx* and a negative result with the corresponding *wzy* PCR. Sequencing could differentiate individual serotypes (23A, 23F and 23B) except mcst 23F-23A and 23A-23F. Mcst 23F-c, 23A-23F and 23F-23A have apparently arisen by recombination between 23F, 23A and/or others, producing sequences in the *cpsA-cpsB* regions that are quite different from their parental types.

Serogroups 33 and 37

Mct 33A and 33F-g share identical *cpsA-cpsB* sequences and that of 33B is similar; 37 and 33F-g cluster together, as do 33B and 33F-q (Figure 65). The 33F/37-specific *wzx* PCR amplified 37, 33F, 33A and 33B, indicating similarities at that site, although sequencing showed clear differences between 33B and the others. The 33F/37-specific *wzy* PCR amplified 37, 33F and 33A but not 33B. Thus, mct 33B was identified on the basis of a positive result with 33F/37-specific primers targeting *wzx* and a negative result with the corresponding *wzy* PCR.

Other serogroups

Despite antigenic similarities that determine their membership of the same serogroup, mct 9N and 9V appear to be genetically distant, on the basis of significant differences between their *cpsA-cpsB* sequences and the fact that 9V-specific PCR did not amplify 9N.

Similarly, mct 19F and 19A had quite different *cpsA-cpsB* region sequences and separated into different clusters. 19F-specific PCR did not amplify 19A and vice versa. There were differences between mct 19F, 19A, 19B, 19C in *wzx* and *wzy* sequences (except *wzy* sequence of 19C was not available in GenBank), but they formed two groups - 19F, 19A and 19B, 19C.

Mct 7F and 7C separated into different clusters based on *cpsA-cpsB* sequences, as did 11A and 11B (Figure 65). Mct 15B and 15C had similar *cpsA-cpsB* sequences and clustered together, except for mcst 15B-22F. Mct 17F (including mcst 17F-c and 17F-35B) and 17A were clustered together. Mct 22F and 22A can be distinguished on the basis of a single but very stable heterogeneity site. Mct 35F and 35B are closely related based on similar *cpsA-cpsB* sequences.

Mixed culture

One clinical isolate identified as serotype 9/14 using antisera was positive in 9V- and 14-specific PCR (targeting both *wzx* and *wzy*), but was identified as mct 9V by sequencing. The isolate was subcultured and 16 individual colonies were rested. All 16 colonies were positive in both mct 9V-specific and negative in both mct 14-specific PCR assays and were identified as mct 9V by sequencing. The serotype of the original isolate was rechecked and the results (mixed serotype 9/14) were as before. It was therefore assumed that the original isolate was a mixture, predominantly of serotype/mct 9V with a minor component of serotype/mct 14.

Comparison of serotype identification results between MCT and CS

After CS and MCT had been completed, the results were compared. Initial results were discrepant for 29 isolates; repeat serotyping and/or correction of clerical errors resolved all but five discrepancies. Final results correlated between CS and MCT methods for all isolates of 38 serotypes (318 isolates), 20 of 25 of another three serotypes and all five nonserotypable isolates (total 343 isolates). In addition, there were 46 isolates belonging to pairs of serotypes whose members could not be distinguished from each other by MCT but all were assigned to the pair that included the serotype to which they had been assigned by CS. These results were classified as consistent.

The five discrepant results were: one isolate of serotype 6A was identified as mcst 6B-q, two isolates of serotype 15B were identified as mct 22F and two isolates of serotype 17F as mct 35B.

Algorithm for serotype assignment of *S. pneumoniae* by MCT

An algorithm for practical use of the MCT method for the identification of *S. pneumoniae* serotypes is shown in Table 6.

DISCUSSION

Sequences of 16 *cps* gene clusters showed that all have the same four genes at their 5' ends - *cpsA (wzg)-cpsB (wzh)-cpsC (wzd)-cpsD (wze)* - which are the sites for recombination events that generate new forms of capsular polysaccharide. The sequences for different serotypes can be divided into two classes and show evidence of interesting recombination patterns.

Table 6. Algorithm for *S. pneumoniae* molecular capsular type (mct) identification by sequencing and mct-specific PCR.

Amplification primer pairs*	PCR product size (base pairs)	Interpretation
	***S. pneumoniae* identification primer pairs**	
P1/P2	864	*S. pneumoniae*
	***S. pneumoniae* mct identification by sequencing**	
CpsS1/cpsA3 (for most mct)	1001	1. Purification PCR amplicons
or	or	2. Sequencing PCR amplicons
cpsS1/cpsA1+ cpsS3/cpsA2 (for mct 38/25F and some nontypable isolates)	520+503	3. Using programmes (Pileup & Pretty or Ednadist & Ekitsch etc.) in ANGIS to analyse sequences to identify mct/mcst
		4. Refer to Figure 1/Table 4 to identify/confirm mct/mcst.
	***S. pneumoniae* mct identification by mct-specific PCR**	
See Table 2 for primer sequences* and Table 3 for specificity and amplicon lengths of primer pairs. Only selected mct and isolates need to be identified using the full testing algorithm.		

The study of 51 serotypes, of which 40 were represented by more than one isolate, showed that the *cpsA-cpsB* sequences for the same serotypes were generally stable or could be consistently divided into a small number of subtypes. This shows that sequence patterns in this region can be used to identify different serotypes/serosubtypes.

It has been shown previously that PCR-RFLP based on the *cpsA-cpsB* region can predict *S. pneumoniae* serotypes (Lawrence et al., 2000). However, the method generates a long amplicon (1.8kbp), requires the use of three restriction enzymes and special equipment and has limited discriminatory ability.

The present inventors identified 376 sequence heterogeneity sites, in the *cpsA-cpsB* region, among the 51 serotypes studied (Table 4, Figure 2), which allowed a practical MCT assay based on sequencing to be developed. Several pairs of primers were designed to amplify a 1001 bp segment within the *cpsA-cpsB* region, based on the following considerations. The primers formed amplicons from virtually all, *S. pneumoniae* isolates (>99% of those examined); the amplicon is small enough to be amplified using normal PCR protocols; the region of interest (800bp) can be sequenced using a single reaction and the method is objective. The target included most of the variable sites (bp 951 to 1747), providing maximum discrimination between closely related serotypes (e.g. members of serogroups 33 and 37 that could not be distinguished by serotype-specific PCR).

Some of the 376 heterogeneity sites in the *cpsA-cpsB* region were specific for individual mct or mcst (Table 4, Figure 2), while others were shared between several. Based on these patterns; plus PCR and selective sequencing of type-specific regions of *wzx* and *wzy*, most of the 51 serotypes represented among our isolates could be distinguished and further divide them into a total of 71 mct and mcst, with the aid of sequence analysis software. The final CS and MCT results correlated for 343 isolates of 389 (88%) for which results for both methods were available, including five that were nontypable by either method. For 46 isolates belonging to five serotype pairs, members of which could not be distinguished by sequencing, results were classified as consistent leaving unresolved discrepancies between methods for only five (1.2%) isolates.

Sequence analysis of the *cps* gene clusters of 16 serotypes showed that *wzy* (capsular polysccharide polymerase gene) and *wzx* (capsular polysccharide flippase gene) are highly variable, making them suitable targets for direct serotype identification by PCR. The present inventors designed serotype-specific PCR primers for these serotypes, targeting *wzx* and *wzy* and, for serotype 3, which has no *wzy* and *wzx* genes, targeting *orf2 (wze)-cap3A- cap3B* (Arrecubieta et al., 1996). It was found that

presumed serotype-specific primers for 6A, 18C, 23F and 33F/37 were not serotype-specific, but amplified other related serotypes. To improve the MCT methods, portions of the *wzy* and *wzx* genes of serotypes within these groups were sequenced, which allowed mct and, in some cases, mcst to be distinguished within these serogroups and demonstrate relationships between them.

The present inventors have recognized that the large number of pneumococcal serotypes would make it impractical to use serotype-specific PCR for all of them. Nevertheless, *wzy* and *wzx* PCR can be used to resolve discrepancies between CS and *cpsA-cpsB* region sequencing assays e.g. for mcst 10A-23F and 23A-23F. Moreover, the use of two target regions in the *cps* gene cluster helps to clarify the relationships between mcst that have apparently arisen by recombination. Mct-specific primers were evaluated using three reference panels, which had been characterised by CS and used to identify clinical isolates of unknown cs. By PCR alone, 239 (61%) of our 394 clinical isolates were assigned to a serotype or serogroup (Table 5). This method can be extended to other mct, when additional *wzx* and *wzy* sequences are available.

In some circumstances, sequencing of the *cpsA-cpsB* region may be more practical than type-specific PCR. For most serotypes only a single method and fewer primers (cpsS1/*cpsA*3-for most serotypes/isolates) are needed.

Previous studies have shown that serotypes included in 23-valent polysaccharide and 11-, 9-, 7-valent protein conjugate vaccines are those most frequently isolated from normally sterile sites (CSF, blood) (Colman et al., 1998; Huebner et al., 2000). Among 173 consecutive pneumococcal "sterile site" isolates from adults in the CIDM diagnostic laboratory, over a 2.5-year period, correlation between the mct and cs was good (171/173 CIDM isolates were correctly identified). The exceptions were two cs 15B isolates that were identified as mct 22F. Five serotypes (4, 14, 19F, 23F, 9V – covered by all pneumococcal vaccines) accounted for 57% of isolates.

Five of 394 isolates studied were nontypable by both CS and MCT (Barker et al., 1999). Isolates may be nonserotypable because of decreased type-specific-antigen synthesis, nonencapsulated phase variation or insertion or mutation of genes of *cps* gene clusters. Failure to type them by MCT reflects the fact that the sequence database is still incomplete, although the target regions of two of the five nonserotypable isolates have been sequenced.

In summary, the present inventors have developed a MCT system for *S. pneumoniae*, which is reproducible, can be performed by any laboratory with access to PCR/sequencing and does not require large panels of expensive serotype-specific antisera. Work on an international collection of isolates in our reference panels

demonstrated a strong correlation between the *cpsA-cpsB* sequence and cs. Heterogeneity in a relatively short sequence (800bp) in this region, supplemented by type-specific PCR targeting *wzx* and *wzy*, correctly predicted the serotype of most unknown isolates belonging to 51 serotypes. These novel MCT methods provide comprehensive strain identification that will be useful for epidemiological studies that will be needed to monitor serotype distribution and detect serotype switching, if any, among *S. pneumoniae* isolates before and following introduction and widespread use of conjugate vaccines.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

All publications discussed above are incorporated herein in their entirety.

Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is solely for the purpose of providing a context for the present invention. It is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed before the priority date of each claim of this application.

Dated this tenth day of April 2003

Western Sydney Area Health Service
Patent Attorneys for the Applicant:

F B RICE & CO

REFERENCES

Arai, S. et al. (2001) Microbiol. Immunol. 45;159-62.

Arrecubieta, C. et al. (1996) J. Exp. Med. 184;449-55.

Barker, J.H. et al. (1999) J. Clin. Microbiol. 37;4039-41.

Coffey, T.J. et al. (1998) Mol. Microbiol. 27;73-83.

Colman, G. et al. (1998). J. Med. Microbiol. 47;17-27.

Dunne, W.M., Jr. (2001) J. Clin. Microbiol. 39;1791-1795.

Hausdorff, W.P. et al. (2001) Lancet 357;950-2.

Henrichsen, J. (1999) Am. J. Med. 107;50S-54S.

Huebner, R.E. et al. (2000) S. Afr. Med. J. 90;1116-21.

Huebner, R.E. et al. (2000) Int. J. Infect. Dis. 4;214-8.

Jiang, S M. et al. (2001) Infect. Immun. 69;1244-55.

Kong, F. et al. (2000) J. Clin. Microbiol. 38;4256-9.

Kong, F. et al. (2002) J. Clin. Microbiol. 40;216-26.

Lalitha, M.K. et al. (1999) J. Clin. Microbiol. 37;263-5.

Lawrence, E.R. et al. (2000) J. Clin. Microbiol. 38;1319-23.

Lipsitch, M. (2001) Am. J. Epidemiol. 154;85-92.

Morrison, K.E. et al. (2000) J. Clin. Microbiol. 38;434-7.

Rubins, J.B. et al. (1999) Infect. Immun. 67;5979-84.

Salo, P. et al. (1995). J. Infect. Dis. 171;479-82.

Schena. M. et al. (1998) Trends Biotechnol. 16;301-6.

Sorensen, U.B. (1993) J.Clin. Microbiol. 31;2097-2100.

Straub, T.M. et al. (2002) Appl. Environ. Microbiol. 68;1817-26.

Vakevainen, M. et al. (2001) J. Infect. Dis. 184;789-93.

van Selm, S. et al. (2002) Microbiology 148;1747-55.

Volokhov, D. et al. (2002) J. Clin. Microbiol. 40;4720-8.

1	gtcaaatctg	tcttgattga	aaacactgcg	gctaaagaag	tacttgaaaa	acaggtcttg
61	gctccatggg	atgctttctg	tgtggaatta	ctataaatat	tttttgcaga	aaaatttaaa
121	attgaaatcg	tataaaaaca	agggaggact	gtataaaaga	cagaaatcct	ttgtttttta
181	taaccaaagt	ttataaactt	tcattcttga	aattcaatta	actttacaaa	ttcccactat
241	taaggagaaa	gaagatgaac	ataaagaagc	gtgtccttag	tgcaggcctg	acttttgcat
301	ctgctttgct	tttagctgct	tgcggccaat	caggttcaga	tacaaaaact	tactcatcaa
361	cctttagtgg	aaatccaact	acatttaact	atctattaga	ctattacgct	gataatatag
421	ttaattgaaa	caagaacaag	acaaaagagc	ctcataaaag	gtattgcaac	ttggtaatac
481	ctttttgagg	tgctttttga	tatgagccca	tgttttctca	ataggattgt	actcaggtga
541	gtagggagga	agaggtaaaa	gtttataccc	aaactcttca	cacaagagtt	ctaacttacc
601	cattctatgg	aatcttgcat	tatccataat	aataaccgat	ggtgtgttta	atgttggtaa
661	gagaaatttc	tgaaaccaag	cttcaaaaaa	gtcgctcgtc	atcgtctctt	cgtaagttat
721	tggagcgatt	aactcaccat	ttgttagacc	tgcaaccaaa	gaaatcctct	gatatcttct
781	tccagatact	ttgcctcttc	ttaactgacc	ttttaatgag	cgaccatatt	ctcgataaaa
841	ataagtatcg	aatcctgttt	catcaatcta	aacaggtgct	aggtgcttta	aactattaaa
901	attcttaaga	aataaggcta	ctttttctgg	gttttgttca	tagtaggtgt	ggttcttttt
961	ttcgagtgta	gcccatagct	ttgagcgcat	agtggatggt	agttggatga	cagccaaatt
1021	cagaagctat	ttcagtcaaa	taagcgtctg	gattgtcagt	aagatagttt	ttaagtctat
1081	ctctatcaac	ttttcttggt	tttgttcctt	ttacttggtg	gtttagctct	cctgtttttct
1141	cttttagctt	taaccagcca	taaatggtat	tacgtgatat	ttggaaaacg	tgtgatgctt
1201	ctgttatact	acctgttcgc	tcacaataag	agagaacttt	tttacgaaaa	tctattgaat
1261	atgccataag	aagattatac	cacattgtgt	actatattag	attgaaacta	gaatagtaca
1321	cctctgcttc	taaaacattg	ttagaaatcg	atttgactgt	cctgaacgat	ttgttctgtt
1381	cttatttcat	tttactatat	ttttgtttcg	cgggaagtct	actaagatac	ttaaagatgc
1441	agatagtaaa	aataaaggtg	tagacattac	cgtaaaaaag	tgatataatc	gtatagtgtt
1501	caatgtatag	gtattaatca	tgagtagacg	ttttaaaaaa	tcacgttcac	agaaagtgaa
1561	gcgaagtgtt	aatatagttt	tgctgactat	ttatttattg	ttagtttgtt	ttttattgtt
1621	cttaatcttt	aagtacaata	tccttgcttt	tagatatctt	aatctagtgg	taactgcgtt
1681	agtcctacta	gttgccttgg	tagggctact	cttgattatc	tataaaaaag	ctgaaaagtt
1741	tactattttt	ctgttggtgt	tctctatcct	tgtcagctct	gtgtcgctct	ttgcagtaca
1801	gcagtttgtt	ggactgacca	atcgtttaaa	tgcgacttct	aattactcag	aatattcaat
1861	cagtgtcgct	gttttagcag	atagtgagat	cgaaaatgtt	acgcaactga	cgagtgtgac
1921	agcaccgact	gggactaata	atgaaaatat	tcagaaatta	ctagctgata	tcaagtcaag
1981	tcagaatacc	gatttgacgg	tcaaccagag	ttcgtcttac	ttggcagctt	acaagagttt
2041	gattgcaggg	gagactaagg	ccattgtcct	aaatagtgtc	tttgaaaaca	tcatcgagtc
2101	agagtatcca	gactacgcat	cgaagataaa	aaagatttat	actaagggat	tcactaaaaa
2161	agtagaagct	cctaagacgt	ctaagagtca	gtctttcaat	atctatgtta	gtggaattga
2221	cacctatggt	cctattagtt	cggtgtcgcg	atcagatgtc	aacatcctga	tgactgtcaa
2281	tcgagatacc	aagaaaatcc	tcttgaccac	aacgccacgt	gatgcctatg	taccaatcgc
2341	agatggtgga	aataatcaaa	aagataaatt	gactcatgcg	ggcatttatg	gagttgattc
2401	gtccattcac	accttagaaa	atctctatgg	agtggatatc	aattactatg	tgcgattgaa
2461	cttcacttcg	tttttgaaat	tgattgattt	gttgggtgga	attgatgttt	ataatgatca
2521	agaatttact	gcccatacga	atggaaagta	ttaccctgca	ggcaatgttc	atcttgattc
2581	agaacaggct	ctcggttttg	ttcgtgagcg	ctactcccta	gcagatggcg	atcgtgaccg
2641	cgggcgccat	caacaaaagg	tgattgtggc	tatccttcaa	aaattaacgt	caaccgaagt
2701	gctgaaaaat	tatagtacga	tcattaatag	cttgcaagat	tctatccaaa	caaatatgcc
2761	acttgagacc	atgataaatt	tggtcaatgc	tcagttagaa	agtggaggga	attataaagt
2821	aaattctcaa	gatttaaaag	ggacaggtcg	gatggatctt	ccttcttatg	caatgccaga
2881	cagtaacctc	tatgtgatgg	aaatagatga	tagtagttta	gctgtagtta	aagcagctat
2941	acaggatgtg	atggagggta	gatgaaatga	tagacatcca	ttcgcatatc	gtttttgatg
3001	tagatgacgg	tcccaagtca	agagaggaaa	gcaaggctct	cttggcagaa	tcctacagac

Continue next page

3061 aggggggtgcg aaccattgtt tctacctctc accgtcgcaa gggcatgttt gaaactccgg
3121 aagagaagat agcagaaaac tttcttcagg ttcgggaaat agctaaggaa gtggcgagtg
3181 acttggtcat tgcttacggg gctgaaattt attacacacc agatgttctg gataagctgg
3241 aaaaaaagcg gattccgacc ctcaatgata gtcgttatgc cttgatagag tttagtatga
3301 acactcctta tcgcgatatt catagcgcct tgagcaagat cttgatgttg ggaattactc
3361 cagtcattgc ccacattgag cgctatgatg ctcttgaaaa taatgaaaaa cgcgttcgag
3421 aactgatcga tatgggctgt tacacgcaag taaatagttc acatgtcctc aaacccaaac
3481 tttttggcga acgttataaa ttcatgaaaa aaagagctca gtatttttta gagcaggatt
3541 tggttcatgt cattgcaagt gatatgcaca atctagacgg tagacctcct catatggcag
3601 aagcatatga ccttgttacc caaaaatacg gagaagcgaa ggctcaggaa ctttttatag
3661 acaatcctcg aaaaattgta atggatcaac taatttagga gaaatgatga aagaacaaaa
3721 tacgatagaa atcgatgtat ttcaattagt taaaagcttg tggaaacgca agctaatgat
3781 tttaatagtg gcacttgtga caggtgcggg ggcttttgca tatagcactt ttattgttaa
3841 gccagaatat acgagtacca cgcgaattta cgtagtgaat cgcaatcaag gagacaagcc
3901 ggggttgaca aatcaggatt tgcaggcagg aacttatctg gtaaaagact accgtgagat
3961 tatcctttcg caggatgttt tggaggaagt tgtttctgat

Figure 1

```
                          1                                                     50
           Serotype 25F   ------a--g --ag----t- -----aat-- ta-t-----g g-c---ag--
            Serotype 38   ------a--g --ag----t- -----aat-- ta-t-----g g-c---ag--
           Serotype 19A   ---------- t----c--t- -g-----t-- ---------- t-------c-
           Serotype 23B   ---------- t----c--t- -g-----t-- ---------- t-------c-
   Serosubtype 6A-6B-q    g--t------ t-------t- -g-----t-- ---------- t---------
      Serosubtype 6B-q    g--t------ t-------t- -g-----t-- ---------- t---------
           Serotype 11B   ---------g t----c--t- -g-----t-- ---------- t---------
         Serotype 11A-q   ---------- ---------- ---------- -a-------- ----------
      Serosubtype 6A-c    ---------- ---------- ---------- ---------- ----------
     Serosubtype 6A-ca    ---------- ---------- ---------- ---------- ----------
       Serosubtype 6A-g   ---------- ---------- ---------- ---------- ----------

  Serosubtype 15A-ca2     ---------- ---------- ---------- ---------- ----------
    Serosubtype 23F-c     g--t------ t--------- ---------- ---------- ----------
           Serotype 18B   ---------- ---------- ---------- ---------- ----------
           Serotype 18C   ---------- ---------- ---------- ---------- ----------
           Serotype 19F   ---------- ---------- ---------- ---------- ----------
           Serotype 18F   ---------- ---------- ---------- ---------- ----------
             Serotype 1   ---------- ---------- ---------- ---------- ----------
           Serotype 18A   ---------- ---------- ---------- ---------- ----------
            Serotype 13   ---------- ---------- ---------- ---------- ----------
            Serotype 20   ---------- ---------- ---------- ---------- ----------
            Serotype 9N   a--t------ t--------- ---------- ---------- ----------
    Serosubtype 15B-c     ---------- ---------- ---------- ---------- --------c-
           Serotype 16F   ---------- ---------- ---------- -a-------- ----------
 Serosubtype 23A-23F      ---------- ---------- ---------- ---------- ----------
 Serosubtype 23F-23A      ---------- ---------- ---------- ---------- ----------
    Serosubtype 15B-q     a--t------ t--------- ---------- ---------- ----------
    Serosubtype 15C-q     a--t------ t--------- ---------- ---------- ----------
   Serosubtype 15C-ca     a--t------ t--------- ---------- ---------- ----------
 Serosubtype 10A-23F      ---------- ---------- ---------- ---------- ----------
    Serosubtype 23F-g     ---------- ---------- ---------- ---------- ----------
     Serosubtype 14-g     ---------- ---------- ---------- ---------- ----------
            Serotype 29   a--t------ t--------- ---------- ---------- ----------
            Serotype 7F   ---------- ---------- ---------- ---------- ----------
     Serosubtype 14-c     ---------- t----c--t- -g-----t-- ---------- t-------c-
      Serosubtype 5-q     ---------- ---------- ---------- ---------- ----------
       Serosubtype 2-g    ---------- ---------- ---------- ---------- ----------

           Serotype 41F   ---------- ---------- ---------- ---------- ----------
            Serotype 31   ---------- ---------- ---------- ---------- ----------
            Serotype 42   ---------- ---------- ---------- ---------- ----------
      Serosubtype 5-c     ---------- ---------- ---------- ---------- ----------
             Serotype 8   ---------- ---------- ---------- ---------- ----------
           Serotype 33B   ---------- ---------- ---------- ---------- ----------
    Serosubtype 33F-q     ---------- ---------- ---------- ---------- ----------
   Serosubtype 11A-nz     ---------- ---------- ---------- ---------- ----------
 Serosubtype 15B-22F      ---------- ---------- ---------- ---------- ----------
           Serotype 22F   ---------- ---------- ---------- ---------- ----------
           Serotype 22A   ---------- ---------- ---------- ---------- ----------
  Serosubtype 15A-ca1     ---------- ---------- ---------- ---------- --------c-
            Serotype 7C   ---------- ---------- ---------- ---------- --------c-
            Serotype 9V   ---------- t--------- ---------- ---------- ----------
     Serosubtype 6B-c     ---------- ---------- ---------- ---------- ----------
            Serotype 21   a--t------ t--------- ---------- ---------- ----------
           Serotype 10F   a--t------ t--------- ---------- ---------- ----------
           Serotype 12F   a--t------ t--------- ---------- ---------- ----------
      Serosubtype 2-q     ---------- ---------- ---------- ---------- ----------
  Serosubtype 6A-6B-g     g--t------ t-------t- -g-----t-- -a-t------ t---------
     Serosubtype 6B-g     g--t------ t-------t- -g-----t-- -a-t------ t---------
   Serosubtype 23A-ca     g--t------ t-------t- -g-----t-- -a-t------ t---------
```

Continue next page

```
          Serotype 37  g--t------ t-------t- -g-----t-- -a-t------ t---------
         Serotype 17A  g--t------ t-------t- -g-----t-- -a-t------ t---------
          Serotype 34  g--t------ t-------t- -g-----t-- -a-t------ t---------
  Serosubtype 17F-35B  g--t------ t-------t- -g-----t-- -a-t------ t---------
         Serotype 35B  g--t------ t-------t- -g-----t-- -a-t------ t---------
         Serotype 33A  g--t------ t-------t- -g-----t-- -a-t------ t---------
    Serosubtype 33F-g  g--t------ t-------t- -g-----t-- -a-t------ t---------
    Serosubtype 17F-c  g--t------ t-------t- -g-----t-- -a-t------ t---------
    Serosubtype 10A-q  g--t------ t-------t- -g-----t-- -a-t------ t---------
           Serotype 4  g--t------ t-------t- -g-----t-- -a-t------ t---------
         Serotype 35F  g--t------ t-------t- -g-----t-- -a-t------ t---------
           Serotype 3  g--t------ t-------t- -g----at-- -a-t------ t---------
            Consensus  TTTCTTGAAA ATGATTGACT TATTGGGAGG GGTAGATGTT CATAATGATC

                       51                                                 100
         Serotype 25F  --------a- -......at- -acgg-aa-g -c-at--t-- t--a---c--
          Serotype 38  --------a- -......at- -acgg-aa-g -c-at--t-- t--a---c--
         Serotype 19A  ----t---a- -t-------- ---------- ---------- t--a------
         Serotype 23B  ----t--ca- -t-------- ---------- ---------- t--a------
 Serosubtype 6A-6B-q   -g-----ca- ------tgc- aat--aaaa- -c-att-ta- t--t------
     Serosubtype 6B-q  -g-----ca- ------tgc- aat--aaaa- -c-att-ta- t--t------
         Serotype 11B  -g-----ca- ------tgc- aat--aaaa- -c-att-ta- t--t------
       Serotype 11A-q  ---------- ---------- ---------- ---------- ----------
     Serosubtype 6A-c  ---------- -a-------- ---------- ---------- ----------
    Serosubtype 6A-ca  ---------- -a-------- ---------- ---------- ----------
     Serosubtype 6A-g  ---------- -a-------- ---------- ---------- ----------

 Serosubtype 15A-ca2   ---------- ---------- ---------- ---------- ----------
    Serosubtype 23F-c  ---------- ---------- ---------- -------g-- ----------
         Serotype 18B  ---------- ---------- ---------- ---------- ----------
         Serotype 18C  ---------- ---------- ---------- ---------- ----------
         Serotype 19F  ---------- ---------- ---------- ---------- ----------
         Serotype 18F  ---------- ---a------ ---------- ---------- ----------
           Serotype 1  ---------- ---------- ---------- ---------- ----------
         Serotype 18A  ---------- --w------- ---------- ---------- --------c-
          Serotype 13  ---------- ---------- ---------t ---------- ----------
          Serotype 20  ---------- ---------- ---------t ---------- ----------
          Serotype 9N  ---------- ---------- a--------- t--------- ----------
    Serosubtype 15B-c  ---------- ---------- ---------- ---------- ----------
         Serotype 16F  ---------- ---------- ---------- ---------- ----------
  Serosubtype 23A-23F  ---------- ---------- ---------- ---------- ----------
  Serosubtype 23F-23A  ---------- ---------- ---------- ---------- ----------
    Serosubtype 15B-q  ---------- ---------- ---------- ---------- ----------
    Serosubtype 15C-q  ---------- ---------- ---------- ---------- ----------
   Serosubtype 15C-ca  ---------- ---------- ---------- ---------- ----------
  Serosubtype 10A-23F  ---------- ---------- ---------- ---------- ----------
    Serosubtype 23F-g  ---------- ---------- ---------- ---------- ----------
     Serosubtype 14-g  ---------- ---------- ---------- ---------- ----------
          Serotype 29  ---------- ---------- ---------- ---------- ----------
          Serotype 7F  ---------- -----c---- ---------- ---------- ----------
     Serosubtype 14-c  ----t---a- -t-------- ---------- ---------- ----------
      Serosubtype 5-q  ---------- ---------- ---------- ---------- ----------
      Serosubtype 2-g  ---------- ---------- ---------- ---------- ----------

         Serotype 41F  ---------- ---------- ---------- ---------- ----------
          Serotype 31  ---------- ---------- ---------- ---------- ----------
          Serotype 42  ---------- ---------- ---------- ---------- ----------
      Serosubtype 5-c  ---------- ---------- ---------- ---------- ----------
           Serotype 8  ---------- ---------- ---------- ---------- ----------
         Serotype 33B  ---------- ---------- ---------- ---------- ----------
    Serosubtype 33F-q  ---------- ---------- ---------- ---------- ----------
   Serosubtype 11A-nz  ---------- ---------- ---------- ---------- ----------
```

Continue next page

Serosubtype 15B-22F	----------	----------	----------	----------	----------
Serotype 22F	----------	----------	----------	----------	----------
Serotype 22A	----------	----------	----------	----------	----------
Serosubtype 15A-ca1	----------	----------	----------	----------	----------
Serotype 7C	----------	----------	----------	----------	----------
Serotype 9V	----------	----------	----------	----------	----------
Serosubtype 6B-c	----------	----------	----------	----------	----------
Serotype 21	----------	----------	----------	----------	----------
Serotype 10F	----------	----------	----------	----------	----------
Serotype 12F	----------	----------	----------	----------	----------
Serosubtype 2-q	----------	----------	----------	----------	----------
Serosubtype 6A-6B-g	--------a-	t--c-atacg	aatggaaagt	---a---t--	---c-----t
Serosubtype 6B-g	--------a-	t--c-atacg	aatggaaagt	---a---t--	---c-----t
Serosubtype 23A-ca	--------a-	t--c-atacg	aatggaaagt	---a---t-c	---c-----t
Serotype 37	---g----a-	t--c-atacg	aatggaaagt	---a---t-a	---c-----t
Serotype 17A	--------a-	t--c-atacg	aatggaaagt	---at--t-c	---c-----t
Serotype 34	--------a-	t--c-atacg	aatggaaagt	---a---t-c	---c-----t
Serosubtype 17F-35B	--------a-	t--c-atacg	aatggaaagt	---a---t-c	---c-----t
Serotype 35B	--------a-	t--c-atacg	aatggaaagt	---a---t-c	---c-----t
Serotype 33A	--------a-	t--c-atacg	aatggaaagt	---a---t-c	---c-----t
Serosubtype 33F-g	--------a-	t--c-atacg	aatggaaagt	---a---t-c	---c-----t
Serosubtype 17F-c	--------a-	t--c-atacg	aatggaaagt	---a---t-c	---c-----t
Serosubtype 10A-q	--------a-	t--c-atacg	aatggaaagt	---a---t-c	---c-----t
Serotype 4	--------a-	t--c-atacg	aatggaaagt	---a---t-c	---c-----t
Serotype 35F	--------a-	t--c-atacg	aatggaaagt	---a---t-c	---c-----t
Serotype 3	--------a-	t--c-atacg	aatggaaagt	---a---t-a	---c-----t
Consensus	AAGAGTTTTC	AGCTCTACAT	GGGAAGTTCC	ATTTCCCAGT	AGGGAATGTC

	101				150
Serotype 25F	------a-t-	--a--tca--	a--t-----c	-----a----	-tc-t--t--
Serotype 38	------a-t-	--a--tca--	a--t-----c	-----a----	-tc-t--t--
Serotype 19A	--ct----t-	----------	---g--a---	--------g-	----t--c--
Serotype 23B	--ct----t-	----------	---g--a---	--------g-	----t--c--
Serosubtype 6A-6B-q	---t----t-	-a--ag----	a--c-c----	--------g-	----t--c--
Serosubtype 6B-q	---t----t-	-a--ag----	a--c-c----	--------g-	----t--c--
Serotype 11B	---t----t-	-a--ag----	a--c-c----	--------g-	-------t--
Serotype 11A-q	----------	-------a--	-t--------	--------g-	-------c--
Serosubtype 6A-c	----------	----------	----------	----------	----------
Serosubtype 6A-ca	----------	----------	----------	----------	----------
Serosubtype 6A-g	----------	----------	----------	----------	----------
Serosubtype 15A-ca2	----------	----------	----------	----------	----------
Serosubtype 23F-c	----------	-------a--	----------	----------	----------
Serotype 18B	----------	----------	----------	-------g--	----------
Serotype 18C	----------	----------	----------	-------g--	----------
Serotype 19F	----------	----------	----------	----------	----------
Serotype 18F	----------	----------	----------	----------	----------
Serotype 1	----------	-------a--	-t--------	----------	----------
Serotype 18A	----------	-------a--	-t--------	----------	----------
Serotype 13	----------	----------	----------	----------	----------
Serotype 20	----------	----------	----------	----------	----------
Serotype 9N	----------	----------	----------	----------	----------
Serosubtype 15B-c	----------	-------a--	----------	----------	----------
Serotype 16F	----------	-------a--	-t--------	--------g-	-------c--
Serosubtype 23A-23F	----------	----------	----------	----------	----------
Serosubtype 23F-23A	----------	----------	----------	----------	----------
Serosubtype 15B-q	----------	-------a--	----------	----------	----------
Serosubtype 15C-q.	---t------	-------a--	----------	----------	----------
Serosubtype 15C-ca	----------	-------a--	----------	----------	----------
Serosubtype 10A-23F	----------	----------	----------	----------	----------
Serosubtype 23F-g	----------	----------	----------	----------	----------
Serosubtype 14-g	----------	----------	----------	----------	----------

Continue next page

```
      Serotype 29  ---------- ---------- ---------- ---------- ----------
      Serotype 7F  ---------- ---------- ---------- ---------- ----------
 Serosubtype 14-c  ---------- -------a-- ---c------ ---------- ----------
  Serosubtype 5-q  ---------- ---------- ---------- ---------- ----------
  Serosubtype 2-g  ---------- -------a-- ---------- ---------- ----------

     Serotype 41F  ---------- -------a-- ---------- ---------- ----------
      Serotype 31  ---------- -------a-- -t-------- ---------- ----------
      Serotype 42  ---------- -------a-- -t-------- ---------- ----------
  Serosubtype 5-c  ---------- -------a-- -t-------- ---------- ----------
       Serotype 8  ---------- ---------- ---------- ---------- ----------
     Serotype 33B  ---------- ---------- ---------- ---------- ----------
Serosubtype 33F-q  ---------- ---------- ---------- ---------- ----------
Serosubtype 11A-nz ---------- ---------- ---------- ---------- ----------
Serosubtype 15B-22F ---------- ---------- ---------- ---------- ----------
     Serotype 22F  ---------- ---------- ---------- ---------- ----------
     Serotype 22A  ---------- ---------- ---------- ---------- ----------
Serosubtype 15A-ca1 ---------- -------a-- ---------- ---------- ----------
      Serotype 7C  ---------- -------a-- ---------- ---------- ----------
      Serotype 9V  ---------- -------a-- ---------- ---------- ----------
 Serosubtype 6B-c  ---------- ---------- ---------- ---------- ----------
      Serotype 21  ---------- ---------- ---------- ---------- ----------
     Serotype 10F  ---------- ---------- ---------- ---------- ----------
     Serotype 12F  ---------- ---------- ---------- ---------- ----------
  Serosubtype 2-q  ---------- ---------- ---c------ ---------- ----------
Serosubtype 6A-6B-g -----t--t- -a--a----- ---c------ --------g- ----------
 Serosubtype 6B-g  -----t--t- -a--a----- ---c------ --------g- ----------
Serosubtype 23A-ca -----t--t- -a--a----- ---c------ --------g- ----------
      Serotype 37  -----t--t- -a--a----- ---c------ --------g- ----------
     Serotype 17A  -----t--t- -a--a----- -a-c------ --------g- -------c--
      Serotype 34  -----t--t- -a--a----- ---c------ --------g- -------c--
Serosubtype 17F-35B -----t--t- -a--a----- ---c------ --------g- ----------
     Serotype 35B  -----t--t- -a--a----- ---c------ --------g- ----------
     Serotype 33A  -----t--t- -a--a----- ---c------ --------g- ----------
Serosubtype 33F-g  -----t--t- -a--a----- ---c------ --------g- ----------
Serosubtype 17F-c  -----t--t- -a--a----- ---c------ --------g- ----------
Serosubtype 10A-q  -----t--t- -a--a----- ---c------ --------g- -------c--
       Serotype 4  -----t--t- -a--a----- ---c------ --------g- -------c--
     Serotype 35F  -----t--t- -a--a----- ---c------ --------g- ----------
       Serotype 3  -----t--t- -a--a----- ---c------ --------g- -------c--
        Consensus  CATCTAGACT CTGAGCAGGC TCTAGGTTTT GTACGTGAAC GCTACTCACT

                   151                                              200
     Serotype 25F  ---a------ --t------- -a--a--g-- t-----g--- ---c-c-ct-
      Serotype 38  ---a------ --t------- -a--a--g-- t-----g--- ---c-c-ct-
     Serotype 19A  ---g-----t --t-a----- -c--g----- t--------- -------c--
     Serotype 23B  ---g-----t --t-a----- -c--g----- t--------- -------c--
Serosubtype 6A-6B-q ---g-----t --t------- ----g----- t--------- ----------
 Serosubtype 6B-q  ---g-----t --t------- ----g----- t--------- ----------
     Serotype 11B  ---a-----t --t------- ----g----- t--------- ----------
    Serotype 11A-q ---a-----c --t------- ---------- ---------- ----------
 Serosubtype 6A-c  ---------- ---------- ---------- ---------- --c-----a-
Serosubtype. 6A-ca ---------- ---------- ---------- ---------- --c-----a-
 Serosubtype 6A-g  ---------- ---------- ---------- ---------- --c-----a-

Serosubtype 15A-ca2 ---------- ---------- ---------- ---------- --c-----a-
Serosubtype 23F-c  ---a-----c --t------- ---------- ---------- --c-----a-
     Serotype 18B  ---------- ---------- ---------- ---------- --c-----a-
     Serotype 18C  ---------- ---------- ---------- ---------- --c-----a-
     Serotype 19F  ---------- ---------- ---------- ---------- --c-----a-
     Serotype 18F  ---------- t--------- ---------- ---------- --c-----a-
       Serotype 1  -a-------- ---------- ---------- ---------- --c-----a-
```

Continue next page

```
           Serotype 18A  ---------- ---------- ---------- t--------- --c-----a-
            Serotype 13  ---------- ---------- ---------- ---------- ----------
            Serotype 20  ---------- ---------- ---------- ---------- ----------
            Serotype 9N  ---------- ---------- ---------- ---------- ----------
     Serosubtype 15B-c   ---------- ---------- ---------- ---------- ----------
           Serotype 16F  ---a-----c --t------- ---------- ---------- ----------
 Serosubtype 23A-23F     ---a-----c --t------- -c--g----- t--------- ----------
 Serosubtype 23F-23A     ---a-----c --t------- -c--g----- t--------- ----------
     Serosubtype 15B-q   ---------- ---------- ---------- ---------- ----------
     Serosubtype 15C-q   ---------- ---------- ---------- ---------- ----------
    Serosubtype 15C-ca   ---------- ---------- ---------- ---------- ----------
 Serosubtype 10A-23F     ---------- ---------- ---------- ---------- --t-----a-
     Serosubtype 23F-g   ---------- ---------- ---------- ---------- --t-----a-
      Serosubtype 14-g   ---------- ---------- ---------- ---------- ----------
            Serotype 29  ---------- ---------- ---------- ---------- ----------
            Serotype 7F  ---------- ---------- ---------- ---------- ----------
      Serosubtype 14-c   ---------- ---------- ---------- ---------- ----------
       Serosubtype 5-q   ---------- ---------- ---------- ---------- ----------
       Serosubtype 2-g   ---------- ---------- -c--g----- t--------- ----------

           Serotype 41F  ---------- ---------- -c--g----- t--------- ----------
            Serotype 31  -a-------- ---------- ---------- ---------- ----------
            Serotype 42  -a-------- ---------- ---------- ---------- ----------
       Serosubtype 5-c   -a-------- ---------- ---------- ---------- ----------
             Serotype 8  ---------- ---------- ---------- ---------- ----------
           Serotype 33B  ---------- --------t- -c--g----- t--------- ----------
     Serosubtype 33F-q   ---------- --------t- -c--g----- t--------- ----------
    Serosubtype 11A-nz   ---------- ---------- ---------- ---------- ----------
 Serosubtype 15B-22F     ---------- ---------- ---------- ---------- ----------
           Serotype 22F  ---------- ---------- ---------- ---------- ----------
           Serotype 22A  ---------- ---------- ---------- ---------- ----------
   Serosubtype 15A-ca1   ---------- ---------- ---------- ---------- ----------
            Serotype 7C  ---------- ---------- ---------- ---------- ----------
            Serotype 9V  ---------- ---------- ---------- ---------- ----------
      Serosubtype 6B-c   ---------- ---------- ---------- ---------- --t-----a-
            Serotype 21  ---------- ---------- ---------- ---------- ----------
           Serotype 10F  ---------- ---------- ---------- ---------- ----------
           Serotype 12F  ---------- ---------- ---------- ---------- ----------
       Serosubtype 2-q   ---------- ---------- -c--g----- t--------- ----------
   Serosubtype 6A-6B-g   ---a-----c --t------- ---------- ---------- --c-----a-
      Serosubtype 6B-g   ---a-----c --t------- ---------- ---------- --c-----a-
    Serosubtype 23A-ca   ---a-----c --t------- -c--g----- t--------- ----------
            Serotype 37  ---a-----c --t------- -c--g----- t--------- ----------
           Serotype 17A  ---a-----c --t------- ---------- ---------- ----------
            Serotype 34  ---a-----c --t------- ---------- ---------- ----------
 Serosubtype 17F-35B     ---a-----c --t------- -c--g----- t--------- ----------
           Serotype 35B  ---a-----c --t------- -c--g----- t--------- ----------
           Serotype 33A  ---a-----c --t------- -c--g----- t--------- ----------
     Serosubtype 33F-g   ---a-----c --t------- -c--g----- t--------- ----------
     Serosubtype 17F-c   ---a-----c --t------- -c--g----- t--------- --------a-
     Serosubtype 10A-q   ---a-----c --t------- -c--g----- t--------- ----------
             Serotype 4  ---a-----c --t------- -c--g---c- t--------- ----------
           Serotype 35F  ---a-----c --t------- ----g----- t--------- --c-------
             Serotype 3  ---a-----c --t------- -c--g----- t--------- ----------
              Consensus  AGCCGATGGA GACCGTGACC GTGGTCGCAA CCAACAAAAG GTGATTGTGG

                         201                                                 250
           Serotype 25F  -a--ta---- ---------c ----ta---t --t------- ----c--caa
            Serotype 38  -a--ta---- ---------c ----ta---t --t------- ----c--caa
           Serotype 19A  -c-------- ---------t --g-g----- -g-------- --------g-
           Serotype 23B  ---------- ---------t --g-g----- -g-------- --------t-
   Serosubtype 6A-6B-q   ---------- ---------t --g------- t--------- ----------
```

Continue next page

```
     Serosubtype 6B-q  ---------- ---------t --g------- t--------- ----------
       Serotype 11B    ---------- ---------t --g------- tg-------- ------c-ga
       Serotype 11A-q  ---------- ---------- ---------- -g-------- ----------
     Serosubtype 6A-c  -a--ta--a- g--g-----t --t--a--g- ttt------- c-------g-
    Serosubtype 6A-ca  -a--ta--a- g--g-----t --t--a--g- ttt------- c-------g-
     Serosubtype 6A-g  -a--ta--a- g--g-----t --t--a--g- ttt------- c-------g-
Serosubtype 15A-ca2    -a--ta---- g--g-----t --t--a--g- ttt------- c-------g-
   Serosubtype 23F-c   -a--ta---- g--g-----t --t--a--g- ttt------- c-------a-
        Serotype 18B   -a--ta---- g--g-----t --t--a--g- ttt------- c-------g-
        Serotype 18C   -a--ta---- g--g-----t --t--a--g- ttt------- c-------g-
        Serotype 19F   -a--ta---- g--g-----t --t--a--g- ttt------- c-------g-
        Serotype 18F   -a--ta---- g--g-----t --t--a--g- ttt------- c-------g-
          Serotype 1   -a--ta---- g--g-----t --t--a--g- ttt------- c-------g-
        Serotype 18A   -g--ta---- g--g-----t --t--a--g- tct-----g- c-------g-
         Serotype 13   ---------- ---------- ---------- ---------- ----------
         Serotype 20   ---------- ---------- ---------- ---------- ----------
         Serotype 9N   ----t----- ---------- ---------- ---------- ----------
   Serosubtype 15B-c   ---------- ---------- ---------- ----t----- ------c---
        Serotype 16F   ---------- ---------- ---------- ---------- ----------
 Serosubtype 23A-23F   ---------- ---------- ---------- ---------- ----------
 Serosubtype 23F-23A   ---------- ---------- ---------- ---------- ----------
   Serosubtype 15B-q   ---------- ---------- ---------- -g-------- ----------
   Serosubtype 15C-q   ---------- ---------- ---------- -g-------- ----------
  Serosubtype 15C-ca   ---------- ---------- ---------- ---------- ----------
 Serosubtype 10A-23F   ---------- ---------- ---------- ---------- ----------
   Serosubtype 23F-g   ---------- ---------- ---------- ---------- ----------
    Serosubtype 14-g   ---------- ---------- ---------- ---------- ----------
         Serotype 29   ---------- ---------- ---------- ---------- ----------
         Serotype 7F   ---------- ---------- ---------- ---------- ----------
    Serosubtype 14-c   ---------- ---------- ---------- ---------- ----------
     Serosubtype 5-q   ---------- ---------- ---------- ---------- ----------
     Serosubtype 2-g   ---------- ---------- ---------- ---------- ----------

        Serotype 41F   ---------- ---------- ---------- ---------- ----------
         Serotype 31   ---------- ---------- ---------- t--------- ----------
         Serotype 42   ---------- ---------- ---------- t--------- ----------
     Serosubtype 5-c   ---------- ---------- ---------- t--------- ------c---
          Serotype 8   ---------- ---------- ---------- t--------- ----------
        Serotype 33B   ---------- ---------- ---------- ---------- ----------
   Serosubtype 33F-q   ---------- ---------- ---------- ---------- ----------
  Serosubtype 11A-nz   ---------- ---------- ---------- ---------- ----------
 Serosubtype 15B-22F   ---------- ---------- ---------- ---------- ----------
        Serotype 22F   ---------- ---------- ---------- ---------- ----------
        Serotype 22A   ---------- ---------- ---------- ---------- ----------
 Serosubtype 15A-ca1   ---------- ---------- ---------- ---------- ----------
         Serotype 7C   ---------- ---------- ---------- ---------- ----------
         Serotype 9V   ---------- ---------- ---------- ---------- ----g-----
    Serosubtype 6B-c   ---------- ---------- ---------- ---------- ----------
         Serotype 21   ---------- ---------- ---------- ---------- ----------
        Serotype 10F   ---------- ---------- ---------- ---------- ----------
        Serotype 12F   ---------- ---------- ---------- ---------- ----------
     Serosubtype 2-q   ---------- ---------- ---------- -g------c- ----------
 Serosubtype 6A-6B-g   -a--ta--a- g--g-----t --t--a--g- ttt------- c-------g-
    Serosubtype 6B-g   -a--ta--a- g--g-----t --t--a--g- ttt------- c-------g-
  Serosubtype 23A-ca   ---------- ---------- ---------- ---------- ------c---
         Serotype 37   ---------- ---------- ---------- ---------- ----------
        Serotype 17A   ---------- ---------- ---------- ---------- ----------
         Serotype 34   ---------- ---------- ---------- ---------- ----------
 Serosubtype 17F-35B   ---------- ---------- ---------- ---------- ------c---
        Serotype 35B   ---------- ---------- ---------- ---------- ------c---
        Serotype 33A   ---------- ---------- ---------- ---------- ------c---
```

Continue next page

```
  Serosubtype 33F-g  ---------- ---------- ---------- ---------- ------c---
  Serosubtype 17F-c  ---------- ---------- ---------- ---------- ------c---
  Serosubtype 10A-q  ---------- ---------- -t-------- -g-------- ----------
         Serotype 4  ---------- ---------- ---------- tg-------- ----------
       Serotype 35F  ---------- ---------- ---------- -g-------- ----------
         Serotype 3  ---------- ---------- ---------- a--------- ----------
          Consensus  CTATCCTTCA AAAATTAACG TCAACCGAAG CACTGAAAAA TTATAGTACG

                     251                                                 300
       Serotype 25F  -----gc-ag -a--ag-g-- ---a--a--- ------c-ct -t-caaca--
        Serotype 38  -----gc-ag -a--ag-g-- ---a--a--- ------c-ct -t-caaca--
       Serotype 19A  ------g--- ---------- ---------- -----c---- ----------
       Serotype 23B  ------g--- ---------- ---------- -----c---- -c--------
Serosubtype 6A-6B-q  ------g--- ---------- ---------- -----c---- ----------
   Serosubtype 6B-q  ------g--- ---------- ---------- -----c---- ----------
       Serotype 11B  ------g--- --------a- ---------- -----c---- -c--------
     Serotype 11A-q  ------g--- ---------- ---------- ---------- ----------
   Serosubtype 6A-c  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
  Serosubtype 6A-ca  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
   Serosubtype 6A-g  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------

Serosubtype 15A-ca2  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
  Serosubtype 23F-c  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
       Serotype 18B  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
       Serotype 18C  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
       Serotype 19F  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
       Serotype 18F  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
         Serotype 1  --tc--c-ag -a-----g-- ----c-t--- --------a- ----------
       Serotype 18A  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
        Serotype 13  ---------- ---------- ---------- ---------- ----------
        Serotype 20  ---------- ---------- ---------- ---------- ----------
        Serotype 9N  ---------- ---------- ---------- ---------- ----------
  Serosubtype 15B-c  ------g--- ---------- ---------- ---------- ----------
       Serotype 16F  ------g--- ---------- ---------- ---------- ----------
Serosubtype 23A-23F  ---------- ---------- ---------- ---------- ----------
Serosubtype 23F-23A  ---------- ---------- ---------- ---------- ----------
  Serosubtype 15B-q  ------g--- ---------- ---------- ---------- ----------
  Serosubtype 15C-q  ------g--- ---------- ---------- ---------- ----------
 Serosubtype 15C-ca  ------g--- ---------- ---------- ---------- ----------
Serosubtype 10A-23F  ------g--- ---------- ---------- ---------- ----------
  Serosubtype 23F-g  ------g--- ---------- ---------- ---------- ----------
   Serosubtype 14-g  ------g--- ---------- ---------- ---------- ----------
        Serotype 29  ---------- ---------- ---------- ---------- ----------
        Serotype 7F  ---------- ---------- ---------- ---------- ----------
   Serosubtype 14-c  ------g--- ---------- ---------- ---------- ----------
    Serosubtype 5-q  ---------- ---------- ---------- ---------- ----------
    Serosubtype 2-g  ---------- ---------- ---------- ------g--- ----------

       Serotype 41F  ---------- ---------- ---------- ------g--- ----------
        Serotype 31  ---------- ---------- ---------- ---------- ----------
        Serotype 42  ---------- ---------- ---------- ---------- ----------
    Serosubtype 5-c  ---------- ---------- ---------- ---------- ----------
         Serotype 8  ---------- ---------- ---------- ---------- ----------
       Serotype 33B  ---------- ---------- ---------- ---------- ----------
  Serosubtype 33F-q  ---------- ---------- ---------- ---------- ----------
 Serosubtype 11A-nz  ---------- ---------- ---------- ---------- ----------
Serosubtype 15B-22F  ---------- ---------- ---------- ---------- ----------
       Serotype 22F  ---------- ---------- ---------- ---------- ----------
       Serotype 22A  ---------- ---------- ---------- ---------- ----------
Serosubtype 15A-cal  ------g--- ---------- ---------- ---------- ----------
        Serotype 7C  ------g--- ---------- ---------- ---------- ----------
        Serotype 9V  ---------- ---------- ---------- ---------- ----------
```

Continue next page

```
   Serosubtype 6B-c  ------g--- ---------- ---------- ---------- ----------
        Serotype 21  ---------- ---------- ---------- ---------- ----------
       Serotype 10F  ---------- ---------- ---------- ---------- ----------
       Serotype 12F  ---------- ---------- ---c------ ------g--- ----------
    Serosubtype 2-q  --t------- ---------- ---------- -----c---- --t-------
Serosubtype 6A-6B-g  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
   Serosubtype 6B-g  --tc--c-ag -a-----g-- ----c-t--- ---------- ----------
 Serosubtype 23A-ca  ---------- ---------- ---------- ---------- ----------
        Serotype 37  ---------- ---------- ---------- ---------- ----------
       Serotype 17A  ---------- ---------- ---------- ---------- ----------
        Serotype 34  ---------- ---------- ---------- ---------- ----------
Serosubtype 17F-35B  ---------- ---------- ---------- ---------- ----------
       Serotype 35B  ---------- ---------- ---------- ---------- ----------
       Serotype 33A  ---------- ---------- ---------- ---------- ----------
  Serosubtype 33F-g  ---------- ---------- ---------- ---------- ----------
  Serosubtype 17F-c  ---------- ---------- ---------- ---------- ----------
  Serosubtype 10A-q  ---------- ---------- ---------- ---------- ----------
         Serotype 4  ---------- ---------- ---------- ---------- ----------
       Serotype 35F  ---------- ---------- ---------- ---------- ----------
         Serotype 3  ------g--- ---------- ---------- ---------- ----------
          Consensus  ATCATTAATA GCTTGCAAGA TTCTATCCAA ACAAATATGC CACTTGAGAC

                     301                                                  350
       Serotype 25F  ---c--gg-c --a-c----- ----a--g-- ---------- ---------a
        Serotype 38  ---c--gg-c --a-c----- ----a--g-- ---------- ---------a
       Serotype 19A  c--------c ---------- ---------- ------t--a -cg--c----
       Serotype 23B  c--------c ---------- ---------- ------t--a -cg--c----
Serosubtype 6A-6B-q  c--------c ---------- ---------- ------t--a -cg--c----
   Serosubtype 6B-q  c--------c ---------- ---------- ------t--a -cg--c----
       Serotype 11B  ---------c ---------- ---------- ------t--a -cg--c----
     Serotype 11A-q  c--------- ---------- ---------- ---------- ----------
   Serosubtype 6A-c  -------g-- --a--g---- -------g-- ------g--- ----------
  Serosubtype 6A-ca  -------g-- --a--g---- -------g-- ---------- ----------
   Serosubtype 6A-g   -------g-- --a--g---- -------g-- ---------- ---------
                                                                          -
Serosubtype 15A-ca2  -------g-- --a--g---- -------g-- ------g--- ----------
  Serosubtype 23F-c  -------g-- --a--g---- -------g-- ------g--- ----------
       Serotype 18B  -------g-- --a--g---- -------g-- ---------- ----------
       Serotype 18C  -------g-- --a--g---- -------g-- ---------- ----------
       Serotype 19F  -------g-- --a--g---- -------g-- ------g--- ----------
       Serotype 18F  -------g-- --a--g---- -------g-- ---------- ----------
         Serotype 1  -------g-- --a--g---- -------g-- ---------- ----------
       Serotype 18A  ------gg-- --a--g---- -c-----g-- ---c------ ----------
        Serotype 13  ---------- ---------- ---------- ---------- ----------
        Serotype 20  ---------- ---------- ---------- ---------- ----------
        Serotype 9N  ---------- ---------- ---------- ------t--a -c---c----
  Serosubtype 15B-c  c------g-- ---------- ---------- ------t--a -c---c----
       Serotype 16F  c--------- ---------- ---------- ------t--a -c---c----
Serosubtype 23A-23F  ---------- ---------- ---------- ---------- ----------
Serosubtype 23F-23A  ---------- ---------- ---------- ---------- ----------
  Serosubtype 15B-q  ---------- ---------- ---------- ---------- ----------
  Serosubtype 15C-q  ---------- ---------- ---------- ---------- ----------
 Serosubtype 15C-ca  ---------- ---------- ---------- ---------- ----------
Serosubtype 10A-23F  ---------- ---------- ---------- ---------- ----------
  Serosubtype 23F-g  ---------- ---------- ---------- ---------- ----------
   Serosubtype 14-g  ---------- ---------- ---------- ---------- ----------
        Serotype 29  ---------a ---------- ---------- ----a----- ----------
        Serotype 7F  ---------- ---------- -------g-- ---------- ----------
   Serosubtype 14-c  ---------- ---------- ---------- ---------- ----------
    Serosubtype 5-q  ---------- ---------- ---------- ---------- ----------
    Serosubtype 2-g   ---------- ---------- ---------- ---------a ---------
                                                                          -
```

Continue next page

```
              Serotype 41F  ---------- ---------- ---------- ---------a ----------
               Serotype 31  -------g-- --a--g---- -------g-- ---------- ----------
               Serotype 42  -------g-- --a--g---- -------g-- ---------- ----------
         Serosubtype 5-c    -------g-- --a--g---- -------g-- ---------- ----------
                Serotype 8  -------g-- --a--g---- -------g-- ---------t ----------
              Serotype 33B  ---------- ---------- ---------- ---------- ----------
       Serosubtype 33F-q    ---------- ---------- ---------- ---------- ----------
      Serosubtype 11A-nz    ---------- -----t---- ---------- ---------- ----------
     Serosubtype 15B-22F    ---------- ---------- ---------- ---------- ----------
              Serotype 22F  ---------- ---------- ---------- ---------- ----------
              Serotype 22A  ---------- ---------- ---------- ---------- ----------
     Serosubtype 15A-ca1    c--------- ---------- ---------- ------a--- ----------
               Serotype 7C  c--------- ---------- ---------- ---------- ----------
               Serotype 9V  c--------- ---------- ---------- ---------- ----------
        Serosubtype 6B-c    ---------- ---------- ---------- ---------- ----------
               Serotype 21  ---------- ---------- ---------- ---------- ----------
              Serotype 10F  ---------- ---------- ---------- ---------- ----------
              Serotype 12F  ---------- ---------- ---------- ---------- ----------
         Serosubtype 2-q    -------g-- --a--g---- -------g-- ---------- ----------
     Serosubtype 6A-6B-g    -------g-- --a--g---- -------g-- ------g--- g---------
        Serosubtype 6B-g    -------g-- --a--g---- -------g-- ------g--- g---------
      Serosubtype 23A-ca    ---------- ---------- ---------- ---------- ----------
               Serotype 37  ---------- ---------- ---------- ---------- ----------
              Serotype 17A  ---------- ---------- ---------- ---------- ----------
               Serotype 34  ---------- ---------- ---------- ---------- ----------
     Serosubtype 17F-35B    ---------- ---------- ---------- ---------- ----------
              Serotype 35B  ---------- ---------- ---------- ---------- ----------
              Serotype 33A  ---------- ---------- ---------- ---------- ----------
       Serosubtype 33F-g    ---------- ---------- ---------- ---------- ----------
       Serosubtype 17F-c    ---------- ---------- ---------- ---------- ----------
       Serosubtype 10A-q    c--------- ---------- ---------- ---------- ----------
                Serotype 4  c--------- ---------- ---------- ---------- ----------
              Serotype 35F  -------g-- --a--g---- -------g-- ---------- --g-------
                Serotype 3  -------g-- --a--g---- -------g-- ---------- ----------
                 Consensus  TATGATAAAT TTGGTCAATG CTCAGTTAGA AAGTGGAGGG AATTATAAAG

                            351                                                400
              Serotype 25F  ---------- ---------- --g-----g- ---------- ----------
               Serotype 38  ---------- ---------- --g-----g- ---------- ----------
              Serotype 19A  -g-----g-- ---c--g--g ----tg--a- ---c------ ------c---
              Serotype 23B  -------g-- ---c--g--g ----gg--a- ---c------ ------c---
     Serosubtype 6A-6B-q    -------g-- ---c--g--g ----gg--a- ---c------ ------c---
        Serosubtype 6B-q    -------g-- ---c--g--g ----gg--a- ---c------ ------c---
              Serotype 11B  -------g-- ---c--g--g ----gg--a- ---c------ ------c---
            Serotype 11A-q  ---------- ---------- --g------- ---c------ ------c---
        Serosubtype 6A-c    -------g-- ---c--g--- ---------- ---------- ----------
       Serosubtype 6A-ca    -------g-- ---c--g--- ---------- ---------- ----------
         Serosubtype 6A-g    -------g-- ---c--g--- ---------- ---------- ---------
                                                                                 -
     Serosubtype 15A-ca2    ---------- ---------- --g------- ---------- ----------
       Serosubtype 23F-c    ---------- ---------- --g------- ---------- ----------
              Serotype 18B  ---------- ---------- --g------- ---------- ----------
              Serotype 18C  ---------- ---------- --g------- ---------- ----------
              Serotype 19F  ---------- ---------- --g------- ---------- ----------
              Serotype 18F  ---------- ---------- --g------- ---------- ----------
                Serotype 1  ---------- ---------- --g------- ---------- ----------
              Serotype 18A  ---------- ---------- --g------- ---------- ------c---
               Serotype 13  ---------- ---------- ---------- ---c------ ----------
               Serotype 20  ---------- ---------- ---------- ---c------ ----------
               Serotype 9N  -g-------- ---------- --g----a-- ---------- ----------
       Serosubtype 15B-c    gg-----g-- ---c--g--- ---------- ---------- ----------
              Serotype 16F  -g-----g-- ---c--g--- ---------- ---------- ----------
```

Continue next page

```
Serosubtype 23A-23F  ---------- ---------- ---------- ---c------ ----------
Serosubtype 23F-23A  ---------- ---------- ---------- ---c------ ----------
  Serosubtype 15B-q  ---------- ---------- ---------- ---c------ ----------
  Serosubtype 15C-q  ---------- ---------- ---------- ---c------ ----------
 Serosubtype 15C-ca  ---------- ---------- ---------- ---c------ ----------
Serosubtype 10A-23F  ---------- ---------- ---------- ---c------ ----------
  Serosubtype 23F-g  ---------- ---------- ---------- ---c------ ----------
   Serosubtype 14-g  ---------- ---------- ---------- ---c------ ----------
        Serotype 29  ---------- ---------- ---------- ---------- ----------
        Serotype 7F  ---------- ---------- --g------- ---------- ----------
   Serosubtype 14-c  ---------- ---------- ---------- ---c------ ----------
    Serosubtype 5-q  ---------- ---------- --g------- ---c------ ----------
    Serosubtype 2-g  ---------- ---------- --g------- ---------- ----------

       Serotype 41F  ---------- ---------- --g------- ---------- ----------
        Serotype 31  ---------- ---------- --g------- ---------- ----------
        Serotype 42  ---------- ---------- --g------- ---------- ----------
    Serosubtype 5-c  ---------- ---------- --g------- ---------- ----------
         Serotype 8  ---------- ---------- --g------- ---------- ----------
       Serotype 33B  ---------- ---------- --g------- ---------- ----------
  Serosubtype 33F-q  ---------- ---------- --g------- ---------- ----------
 Serosubtype 11A-nz  ---------- ---------- --g------- ---c------ ----------
Serosubtype 15B-22F  ---------- ---------- --g------- ---c------ ----------
       Serotype 22F  ---------- ---------- --g------- ---c------ ----------
       Serotype 22A  ---------- ---------- --g------- ---c------ ----------
Serosubtype 15A-ca1  ---------- ---------- ---------- ---------- ----------
        Serotype 7C  ---------- ---------- ---------- ---------- ----------
        Serotype 9V  ---------- ---------- --g------- ---------- ----------
   Serosubtype 6B-c  ---------- ---------- --g------- ---------- ----------
        Serotype 21  ---------- ---------- ---------- ---c------ ----------
       Serotype 10F  ---------- ---------- ---------- ---c------ ----------
       Serotype 12F  ---------- ---------- ---------- ---c------ ----------
    Serosubtype 2-q  ---------- ---------- --g------- -----a---- ----------
Serosubtype 6A-6B-g  ---------- ---------- --g------- ---------- ----------
   Serosubtype 6B-g  ---------- ---------- --g------- ---------- ----------
 Serosubtype 23A-ca  ---------- ---------- ---------- ---c------ ----------
        Serotype 37  ---------- ---------- ---------- ---------- ----------
       Serotype 17A  ---------- ---------- ---------- ---c------ ----------
        Serotype 34  ---------- ---------- ---------- ---c------ ----------
Serosubtype 17F-35B  ---------- ---------- ---------- ---c------ ----------
       Serotype 35B  ---------- ---------- ---------- ---c------ ----------
       Serotype 33A  ---------- ---------- ---------- ---c------ ----------
  Serosubtype 33F-g  ---------- ---------- ---------- ---c------ ----------
  Serosubtype 17F-c  ---------- ---------- ---------- ---c------ ----------
  Serosubtype 10A-q  ---------- ---------- --g------- ---------- ----------
         Serotype 4  ---------- ---------- --g------- ---------- ----------
       Serotype 35F  ---------- ---------- ---------- ---------- ----------
         Serotype 3  ---------- ---------- --g------- ---------- ----------
          Consensus  TAAATTCTCA AGATTTAAAA GGTACAGGTC GGATGGATCT TCCTTCTTAT

                     401                                                  450
       Serotype 25F  --------t- -t---c-gt- g---a----- -----t---- -g-a---cc-
        Serotype 38  --------t- -t---c-gt- g---a----- -----t---- -g-a---cc-
       Serotype 19A  --g------- -t-------- ----a----- -----ta-c- -c-----cc-
       Serotype 23B  --g------- -t-------- ----a----- -----ta-c- -c-----cc-
Serosubtype 6A-6B-q  --g------- -t-------- ---------- -----ta-c- -c-----cc-
   Serosubtype 6B-q  --g------- -t-------- ---------- -----ta-c- -c-----cc-
       Serotype 11B  --g------- -t-------- ----a----- -----ta-c- -c-----cc-
     Serotype 11A-q  --g------- -t-------- ---------- -----ta-c- -c-----cc-
   Serosubtype 6A-c  ---------- ---------- ---------- ---------- ----------
  Serosubtype 6A-ca  ---------- ---------- ---------- ---------- ----------
   Serosubtype 6A-g  ---------- ---------- ---------- ---------- ----------
```

continue next page

Serosubtype 15A-ca2	-----a----	----------	----------	----------	----------
Serosubtype 23F-c	----------	----------	----------	----------	----------
Serotype 18B	----------	----------	----------	----------	----------
Serotype 18C	----------	----------	----------	----------	----------
Serotype 19F	----------	----------	----------	----------	----------
Serotype 18F	----------	----------	----------	----------	----------
Serotype 1	----------	----------	----------	----------	----------
Serotype 18A	----------	----------	----------	----------	----------
Serotype 13	----------	----------	----------	----------	----------
Serotype 20	----------	----------	----------	----------	----------
Serotype 9N	--------t-	----------	----------	----------	----------
Serosubtype 15B-c	----------	----------	----------	----------	----------
Serotype 16F	----------	----------	----------	----------	----------
Serosubtype 23A-23F	----------	----------	----------	----------	----------
Serosubtype 23F-23A	----------	----------	----------	----------	----------
Serosubtype 15B-q	----------	----------	----------	----------	----------
Serosubtype 15C-q	----------	----------	----------	----------	----------
Serosubtype 15C-ca	----------	----------	----------	----------	----------
Serosubtype 10A-23F	----------	----------	----------	----------	----------
Serosubtype 23F-g	----------	----------	----------	----------	----------
Serosubtype 14-g	----------	----------	----------	----------	----------
Serotype 29	----------	----------	----------	----------	----------
Serotype 7F	----------	----------	----------	----------	----c-----
Serosubtype 14-c	----------	----------	----------	----------	----------
Serosubtype 5-q	----------	----------	-------t--	----------	----------
Serosubtype 2-g	----------	----------	----------	----------	----------
Serotype 41F	----------	----------	----------	----------	----------
Serotype 31	----------	----------	----------	----------	----------
Serotype 42	----------	----------	----------	----------	----------
Serosubtype 5-c	----------	----------	----------	----------	----------
Serotype 8	----------	----------	----a-----	----------	----------
Serotype 33B	----------	----------	----------	----------	----------
Serosubtype 33F-q	----------	----------	----------	----------	----------
Serosubtype 11A-nz	----------	----------	----------	----------	----------
Serosubtype 15B-22F	----------	----------	-------t--	----------	----------
Serotype 22F	----------	----------	-------t--	----------	----------
Serotype 22A	----------	----------	-------t--	----------	----------
Serosubtype 15A-ca1	----------	----------	----------	----------	----------
Serotype 7C	----------	----------	----------	----------	----------
Serotype 9V	----------	----------	-------c--	----------	----------
Serosubtype 6B-c	----------	----------	----------	----------	----------
Serotype 21	----------	----------	-------c--	----------	----------
Serotype 10F	----------	----------	----------	----------	----------
Serotype 12F	----------	----------	----------	----------	----------
Serosubtype 2-q	---g------	----------	----------	----------	----------
Serosubtype 6A-6B-g	----------	----------	----------	----------	----------
Serosubtype 6B-g	----------	----------	----------	----------	----------
Serosubtype 23A-ca	----------	----------	----------	----------	----------
Serotype 37	----------	----------	----------	----------	----------
Serotype 17A	----------	----------	----------	----------	----------
Serotype 34	----------	----------	----------	----------	----------
Serosubtype 17F-35B	----------	----------	----------	----------	----------
Serotype 35B	----------	----------	----------	----------	----------
Serotype 33A	----------	----------	----------	----------	----------
Serosubtype 33F-g	----------	----------	----------	----------	----------
Serosubtype 17F-c	----------	----------	----------	----------	----------
Serosubtype 10A-q	----------	----------	----------	----------	----------
Serotype 4	----------	----------	----------	----------	----------
Serotype 35F	----------	----------	----------	----------	----------
Serotype 3	----------	----------	----------	----------	----------
Consensus	GCAATGCCAG	ACAGTAACCT	CTATGTGATG	GAAATAGATG	ATAGTAGTTT

Continue next page

```
                          451                                                500
         Serotype 25F     ct-a-cta-c ----a-aa-- -t----c--- tc-------- --g-a-g---
          Serotype 38     ct-a-cta-c ----a-aa-- -t----c--- tc-------- --g-a-g---
         Serotype 19A     t--atct--c ---a-t---- -t-------- -t-------c ----------
         Serotype 23B     t--atct--c ---a-t---- -t-------- -t-------c ----------
Serosubtype 6A-6B-q       t--atct--c ---a-g---- -t-----c-- -t-------c ----------
     Serosubtype 6B-q     t--atct--c ---a-g---- -t-----c-- -t-------c ----------
         Serotype 11B     t--atct--c ---a-t---- -t-------- -t-------g ----------
       Serotype 11A-q     t--atct--c ---a-t---- -t-----c-- -t-------c ----------
     Serosubtype 6A-c     ---------- ---------- ---------- ---------- ----------
    Serosubtype 6A-ca     ---------- ---------- ---------- ---------- ----------
      Serosubtype 6A-g     ---------- --gca----- ---------- ---------- ---------
                                                                               -

Serosubtype 15A-ca2       ---------- ---------- --g------- ---------- ----------
   Serosubtype 23F-c      ---------- ---------- ---------- ---------- ----------
         Serotype 18B     --------c- ---------- -------c-- ---------- ----------
         Serotype 18C     --------c- ---------- -------c-- ---------- ----------
         Serotype 19F     ---------- ---------- ---------- ---------- ----------
         Serotype 18F     ---------- --------c- ---------- ---------- ----------
           Serotype 1     -a-------- ---------- ---------- ---------- ----a-----
         Serotype 18A     ---------- ---------- ---------- ---------- ----------
          Serotype 13     ---------- ---------- ---------- ---------- ----------
          Serotype 20     ---------- ---------- ---------- ---------- ----------
          Serotype 9N     ---------- ---------- ---------- ---------- ----------
    Serosubtype 15B-c     ---------- ---------- ---------- ---------- ----------
         Serotype 16F     ---------- ---------- ---------- ---------- ----------
Serosubtype 23A-23F       ---------- ---------- ---------- ---------- ----a-----
Serosubtype 23F-23A       ---------- ---------- ---------- ---------- ----a-----
    Serosubtype 15B-q     ---------- ---------- ---------- ---------- ----------
    Serosubtype 15C-q     ---------- ---------- ---------- ---------- ----------
  Serosubtype 15C-ca      ---------- ---------- ---------- ---------- ----------
Serosubtype 10A-23F       ---------- ---------- ---------- ---------- ----------
    Serosubtype 23F-g     ---------- ---------- ---------- ---------- ----------
     Serosubtype 14-g     ---------- ---------- ---------- ---------- ----------
          Serotype 29     ---------- ---------- ---------- ---------- ----------
          Serotype 7F     ---------- ---------- ---------- ---------- ----------
     Serosubtype 14-c     ---------- ---------- ---------- ---------- ----------
      Serosubtype 5-q     ---------- ---------- ---------- ---------- ----------
       Serosubtype 2-g     ---------- ---------- ---------- ---------- ---------
                                                                               -

         Serotype 41F     ---------- ---------- ---------- ---------- ----------
          Serotype 31     ---------- ---------- ---------- ---------- ----------
          Serotype 42     ---------- ---------- ---------- ---------- ----------
      Serosubtype 5-c     ---------- ---------- ---------- ---------- ----------
           Serotype 8     ---------- ---------- ---------- ---------- ----------
         Serotype 33B     ---------- ---------- ---------- ---------- ----------
    Serosubtype 33F-q     ---------- ---------- ---------- ---------- ----------
   Serosubtype 11A-nz     ---------- ---------- ---------- ---------- ----------
Serosubtype 15B-22F       ---------- ---------- ---------- ---------- ----------
         Serotype 22F     ---------- ---------- ---------- ---------- ----------
         Serotype 22A     ---------- ---------- ---------- ---------- ----------
Serosubtype 15A-ca1       ---------- ---------- --g------- ---------- ----------
          Serotype 7C     ---------- ---------- ---------- ---------- ----------
          Serotype 9V     ---------- ---------- ---------- ---------- ----------
     Serosubtype 6B-c     ---------- ---------- ---------- ---------- ----------
          Serotype 21     ---------- ---------- ---------- ---------- ----------
         Serotype 10F     ---------- ---------- ---------- ---------- ----------
         Serotype 12F     ---------- ---------- ---------- ---------- ------g---
      Serosubtype 2-q     ---------- ---------- ---------- ---------- ----------
Serosubtype 6A-6B-g       ---------- ---------- ---------- ---------- ----------
     Serosubtype 6B-g     ---------- ---------- ---------- ---------- ----------
  Serosubtype 23A-ca      ---------- ---------- ---------- ---------- ----a-----
```

Continue next page

```
          Serotype 37   ---------- ---------- ---------- ---------- ----------
         Serotype 17A   ---------- ---------- ---------- ---------- ----------
          Serotype 34   ---------- ---------- ---------- ---------- ----------
  Serosubtype 17F-35B   ---------- ---------- ---------- ---------- ----------
         Serotype 35B   ---------- ---------- ---------- ---------- ----------
         Serotype 33A   ---------- ---------- ---------- ---------- ----------
    Serosubtype 33F-g   ---------- ---------- ---------- ---------- ----------
    Serosubtype 17F-c   ---------- ---------- ---------- ---------- ----------
    Serosubtype 10A-q   ---------- ---------- ---------- ---------- ----------
           Serotype 4   ---------- ---------- ---------- ---------- ----------
         Serotype 35F   ---------- ---------- ---------- ---------- ----------
           Serotype 3   ---------- ---------- ---------- ----a-a--- ----------
            Consensus   AGCTGTAGTT AAAGCAGCTA TACAGGATGT GATGGAGGGT AGATGAAATG

                        501                                                550
         Serotype 25F   -----tg-t- ----t--c-- ta-------- --g------- -g--------
          Serotype 38   -----tg-t- ----t--c-- ta-------- --g------- -g--------
         Serotype 19A   --t--t--t- -------c-- ---c------ ---------- ----------
         Serotype 23B   --t--t--t- -------c-- t--c------ ---------- ----------
  Serosubtype 6A-6B-q   --t--t--t- -------c-- t--c------ ---------- ----------
     Serosubtype 6B-q   --t--t--t- -------c-- t--c------ ---------- ----------
         Serotype 11B   --t--t--t- -------c-- t--c------ ---------- ----------
       Serotype 11A-q   --t--t--t- -------c-- t--c------ ---------- ----------
     Serosubtype 6A-c   ---------- ---------- ---------- ---------- ----------
    Serosubtype 6A-ca   ---------- ---------- ---------- ---------- ----------
     Serosubtype 6A-g   ---------- ---------- ---------- ---------- -c-------
                                                                             -

  Serosubtype 15A-ca2   ------g--- ---------- ---------- ---------- ----------
    Serosubtype 23F-c   ---------- ---------- ---------- ---------- ----------
         Serotype 18B   -----t---- ---------- ---------- ---------- ----------
         Serotype 18C   -----t---- ---------- ---------- ---------- ----------
         Serotype 19F   ---------- ---------- ---------- ---------- ----------
         Serotype 18F   -----t---- ---------- ---------- ---------- ----------
           Serotype 1   ---------- -------c-- ---c------ ---------- ----------
         Serotype 18A   ---------- ---------- ---c------ ---------- ----------
          Serotype 13   ---------- ---------- ---------- ---------- ----------
          Serotype 20   ---------- ---------- ---------- ---------- ----------
          Serotype 9N   ---------- ---------- ---------- ---------- ----------
    Serosubtype 15B-c   ---------- ---------- ---------- ---------- -c--------
         Serotype 16F   ---------- ---------- ---------- ---------- -------a--
  Serosubtype 23A-23F   ---------- ---------- ---------- ---------- ----------
  Serosubtype 23F-23A   ---------- ---------- ---------- ---------- ----------
    Serosubtype 15B-q   ---------- ---------- ---------- ---------- ----------
    Serosubtype 15C-q   ---------- ---------- ---------- ---------- ----------
   Serosubtype 15C-ca   ---------- ---------- ---------- ---------- ----------
  Serosubtype 10A-23F   ---------- ---------- ---------- ---------- ----------
    Serosubtype 23F-g   ---------- ---------- ---------- ---------- ----------
     Serosubtype 14-g   ---------- ---------- ---------- ---------- ----------
          Serotype 29   ---------- ---------- ---------- ---------- ----------
          Serotype 7F   -----t---- ---------- ---------- ---------- ----------
     Serosubtype 14-c   ---------- ---------- ---------- ---------- ----------
      Serosubtype 5-q   ---------- ---------- ---------- ---------- ----------
      Serosubtype 2-g   ------g--- ---------- ---------- ---a------ ---------
                                                                             -

         Serotype 41F   ------g--- ---------- ---------- ---------- ----------
          Serotype 31   ---------- ---------- ---------- ---------- ----------
          Serotype 42   ---------- ---------- ---------- ---------- ----------
      Serosubtype 5-c   ---------- ---------- ---------- ---------- ----------
           Serotype 8   ---------- ---------- ---------- ---------- ----------
         Serotype 33B   ---------- -------c-- ---c------ ---------- ----------
    Serosubtype 33F-q   ---------- -------c-- ---c------ ---------- ----------
   Serosubtype 11A-nz   ---------- -------c-- ---c------ ---------- ----------
```

Continue next page

Serosubtype 15B-22F	----------	----------	----------	----------	----------
Serotype 22F	----------	----------	----------	----------	----------
Serotype 22A	----------	----------	----------	----------	----------
Serosubtype 15A-ca1	------g---	----------	----------	----------	----------
Serotype 7C	----------	----------	----------	----------	----------
Serotype 9V	----------	----------	----------	----------	----------
Serosubtype 6B-c	----------	----------	----------	----------	----------
Serotype 21	----------	----------	----------	----------	-------a--
Serotype 10F	----------	----------	----------	----------	----------
Serotype 12F	----------	----------	----------	----------	----------
Serosubtype 2-q	----------	----------	----------	----------	----------
Serosubtype 6A-6B-g	----------	----------	----------	----------	----------
Serosubtype 6B-g	----------	----------	----------	----------	----------
Serosubtype 23A-ca	----------	----------	----------	----------	----------
Serotype 37	----------	-------c--	----------	----------	----------
Serotype 17A	----------	-------c--	---c------	----------	----------
Serotype 34	----------	-------c--	---c------	----------	----------
Serosubtype 17F-35B	----------	----------	----------	----------	----------
Serotype 35B	----------	----------	----------	----------	----------
Serotype 33A	----------	----------	----------	----------	----------
Serosubtype 33F-g	----------	----------	----------	----------	----------
Serosubtype 17F-c	----------	----------	----------	----------	----------
Serosubtype 10A-q	----------	----------	----------	----------	----------
Serotype 4	----------	----------	----------	----------	----------
Serotype 35F	-----t----	-------c--	---c------	----------	----------
Serotype 3	----------	----------	----------	----------	----------
Consensus	ATAGACATCC	ATTCGCATAT	CGTTTTTGAT	GTAGATGACG	GTCCCAAGTC

	551				600
Serotype 25F	c-t---a--t	--t-ga---t	-g---tt---	-g----tgat	--a--aa-t-
Serotype 38	c-t---a--t	--t-ga---t	-g---tt---	-g----tgat	--a--aa-t-
Serotype 19A	----------	----------	----------	----------	----------
Serotype 23B	----------	--t-------	-----a----	-g--------	----------
Serosubtype 6A-6B-q	----------	--t-------	-----a----	-g--------	----------
Serosubtype 6B-q	----------	--t-------	-----a----	-g--------	----------
Serotype 11B	----------	--t-------	-----a----	-g--------	----------
Serotype 11A-q	----------	----------	-----a----	-g--------	----------
Serosubtype 6A-c	----------	----------	----------	-g-------a	----------
Serosubtype 6A-ca	----------	----------	----------	-g-------a	----------
Serosubtype 6A-g	----------	----------	----------	-g-------a	----------
Serosubtype 15A-ca2	----------	----------	----------	----------	----------
Serosubtype 23F-c	----------	----------	----------	----------	----------
Serotype 18B	----------	----------	----------	-g--------	----------
Serotype 18C	----------	----------	----------	-g--------	----------
Serotype 19F	----------	----------	----------	----------	----------
Serotype 18F	----------	----------	----------	-g--------	----------
Serotype 1	----------	----------	----------	----------	---------a
Serotype 18A	----------	--t-------	----------	-g-------a	----------
Serotype 13	--t-------	--------a-	----aag---	-g-t--t-at	--a--a--t-
Serotype 20	--t-------	--------a-	----aag---	-g-t--t-at	--a--a--t-
Serotype 9N	--t-------	--------a-	----aag---	-g-t--t-at	--a--a--t-
Serosubtype 15B-c	----------	----------	----------	----------	----------
Serotype 16F	----------	----------	----------	-g-------a	----------
Serosubtype 23A-23F	----------	----------	-----a----	----------	----------
Serosubtype 23F-23A	----------	----------	-----a----	----------	----------
Serosubtype 15B-q	----------	----------	-----a----	----------	----------
Serosubtype 15C-q	----------	----------	-----a----	----------	----------
Serosubtype 15C-ca	----------	----------	-----a----	----------	----------
Serosubtype 10A-23F	----------	----------	-----a----	----------	----------
Serosubtype 23F-g	----------	----------	-----a----	----------	----------
Serosubtype 14-g	----------	----------	-----at---	----------	----------

Continue next page

```
         Serotype 29  ---------- ---------- ---------- ---------- ----------
          Serotype 7F  ---------- ---------- ------a--- ---------- ----------
     Serosubtype 14-c  ---------- ---------- -----at--- ---------- ----------
      Serosubtype 5-q  ---------- ---------- ---------- ---------- --a-----a-
      Serosubtype 2-g  ---------- ---------- ---------- ---------- ----------
                                                                            -

         Serotype 41F  ---------- ---------- ---------- ---------- ----------
          Serotype 31  ---------- ---------- ---------- ---------- ----------
          Serotype 42  ---------- ---------- ---------- ---------- ----------
      Serosubtype 5-c  ---------- ---------- ---------- ---------- ----------
           Serotype 8  ---------- ---------- ---------- ---------- ----------
         Serotype 33B  ---------- ---------- -------g-- ---------a ----------
    Serosubtype 33F-q  ---------- ---------- ---------- ---------a ----------
   Serosubtype 11A-nz  ---------- ---------- ---------- -g-------a ------a---
  Serosubtype 15B-22F  ---------- ---------- ---------- ---------- --a-----a-
         Serotype 22F  ---------- ---------- ---------- ---------- --a-----a-
         Serotype 22A  ---------- ------a--- ---------- ---------- --a-----a-
  Serosubtype 15A-ca1  ---------- ---------- ---------- ---------- ----------
          Serotype 7C  ---------- ---------- ---------- ---------- ----------
          Serotype 9V  ---------- ---------- ---------- ---------- ----------
     Serosubtype 6B-c  ---------- ---------- ---------- -g-------a ----------
          Serotype 21  ---------- ---------- ---------- ---------a ----------
         Serotype 10F  ---------- ---------- ---------- ---------- ----------
         Serotype 12F  ---------- ---------- ------t--- -g-------a ----------
      Serosubtype 2-q  ---------- ---------- ---------- ---------- ---------a
  Serosubtype 6A-6B-g  ---------- ---------- ---------- -g-------a ----------
     Serosubtype 6B-g  ---------- ---------- ---------- -g-------a ----------
   Serosubtype 23A-ca  ---------- ---------- -----a---- ---------- ----------
          Serotype 37  ---------- ---------- ---------- ---------- ----------
         Serotype 17A  ---g------ ------a--- ---------- -g-------a ----------
          Serotype 34  ---g------ ------a--- ---------- -g-------a ----------
  Serosubtype 17F-35B  ---------- ---------- -----a---- ---------- ---------a
         Serotype 35B  ---------- ---------- -----a---- ---------- ---------a
         Serotype 33A  ---------- ---------- -----a---- ---------- ---------a
    Serosubtype 33F-g  ---------- ---------- -----a---- ---------- ---------a
    Serosubtype 17F-c  ---------- ---------- -----a---- ---------- ----------
    Serosubtype 10A-q  ---------- ---------- ---------- ---------- ----------
           Serotype 4  ---------- ---------- ---------- ---------a ----------
         Serotype 35F  ---g------ ---------- ------t--- ---------a ----------
           Serotype 3  --t------t --------a- ----aag--- -g-t--t-at --a--a--t-
            Consensus  AAGAGAGGAA AGCAAGGCTC TCTTGGCAGA ATCCTACAGG CAGGGGGTGC

                       601                                                650
         Serotype 25F  -g--a----- t--a--a--c -----c--tc -a--t----- ------a--a
          Serotype 38  -g--a----- t--a--a--c -----c--tc -a--t----- ------a--a
         Serotype 19A  ---------- ---------- ---------- ---------- ----------
         Serotype 23B  ---------- ---------- ---------- ---------- ---------a
  Serosubtype 6A-6B-q  ---------- ---------- ---------- ---------- ---------a
     Serosubtype 6B-q  ---------- ---------- ---------- ---------- ---------a
         Serotype 11B  ---------- ---------- ---------- ---------- ---------a
        Serotype 11A-q  ---------- ---------- ---------- ---------- ---------a
     Serosubtype 6A-c  ---------- ---------- ---------- ---------- ----------
    Serosubtype 6A-ca  ---------- ---------- ---------- ---------- ----------
     Serosubtype 6A-g  ---------- ---------- ---------- ---------- ----------
                                                                            -

  Serosubtype 15A-ca2  ---------- ---------- ---------- ---------- ----------
    Serosubtype 23F-c  ---------- ---------- ---------- ---------- ----------
         Serotype 18B  ---------- ---------- ---------- ---------- ----------
         Serotype 18C  ---------- ---------- ---------- ---------- ----------
         Serotype 19F  ---------- ---------- ---------- ---------- ----------
         Serotype 18F  ---------- ---------- ---------- ---------- ----------
           Serotype 1  ---------- ---------- ---------- ---------- ----------
```

Continue next page

```
         Serotype 18A  ---------- ---------- ---------- ---------- ---------a
          Serotype 13  ---tg----- g-----t--g --t-----a- -a--g----- ---------a
          Serotype 20  ---tg----- g-----t--g --t-----a- -a--g----- ---------a
          Serotype 9N  ---tg----- g-----t--g --t-----a- -a--g----- ----------
    Serosubtype 15B-c  ---------- ---------- ---------- ---------- ----------
         Serotype 16F  ---------- ---------- ---------- ---------- ----------
  Serosubtype 23A-23F  ---------- ---------- ---------- ---------- ----------
  Serosubtype 23F-23A  ---------- ---------- ---------- ---------- ----------
    Serosubtype 15B-q  ---------- ---------- ---------- ---------- ----------
    Serosubtype 15C-q  ---------- ---------- ---------- ---------- ----------
   Serosubtype 15C-ca  ---------- ---------- ---------- ---------- ----------
  Serosubtype 10A-23F  ---------- ---------- ---------- ---------- ----------
    Serosubtype 23F-g  ---------- ---------- ---------- ---------- ----------
     Serosubtype 14-g  ---------- ---------- ---------- ---------- ----------
          Serotype 29  ---------- ---------- ---------- ---------- ----------
          Serotype 7F  ---------- ---------- ---------- ---------- ---------a
     Serosubtype 14-c  ---------- --g------- ---------- ---------- ----------
      Serosubtype 5-q  ---t------ ---------- ---------- ---------- ----------
       Serosubtype 2-g  ---------- ---------- ---------- ----t----- ----------
                                                                             -

         Serotype 41F  ---------- ---------- ---------- ----t----- ----------
          Serotype 31  ----t----- ---------- ---------- ----t----- ----------
          Serotype 42  ----t----- ---------- ---------- ----t----- ----------
      Serosubtype 5-c  ----t----- ---------- ---------- ----t----- ----------
           Serotype 8  ---------- ---------- ---------- ----t----- ----------
         Serotype 33B  ---------- ---------- ---------- ---------- ----------
    Serosubtype 33F-q  ---------- ---------- ---------- ---------- ----------
   Serosubtype 11A-nz  ---------- ---------- ---------- ---------- ----------
  Serosubtype 15B-22F  ---t------ ---------- ---------- ---------- ----------
         Serotype 22F  ---t------ ---------- ---------- ---------- ----------
         Serotype 22A  ---t------ ---------- ---------- ---------- ----------
  Serosubtype 15A-ca1  ---------- ---------- ---------- ---------- ----------
          Serotype 7C  ---------- ---------- ---------- ---------- ----------
          Serotype 9V  ---------- ---------- ---------- ---------- ----------
     Serosubtype 6B-c  ---------- ---------- ---------- ---------- ----------
          Serotype 21  ---------- ---------- --------t- ---------- ----------
         Serotype 10F  ---------- ---------- ---------- ---------- ---------a
         Serotype 12F  ---------- ---------- ---------- ----t----- ----------
      Serosubtype 2-q  ---------- ---------- ---------- ---------- ----------
   Serosubtype 6A-6B-g  ---------- ---------- ---------- ---------- ----------
     Serosubtype 6B-g  ---------- ---------- ---------- ---------- ----------
   Serosubtype 23A-ca  ---------- ---------- ---------- ---------- ----------
          Serotype 37  ---------- ---------- ---------- ---------- ----------
         Serotype 17A  ---------- ---------- ---------- ---------- ----------
          Serotype 34  ---------- ---------- ---------- ---------- ----------
  Serosubtype 17F-35B  ---------- ---------- ---------- ---------- ----------
         Serotype 35B  ---------- ---------- ---------- ---------- ----------
         Serotype 33A  ---------- ---------- ---------- ---------- ----------
    Serosubtype 33F-g  ---------- ---------- ---------- ---------- ----------
    Serosubtype 17F-c  ---------- ---------- ---------- ---------- ----------
    Serosubtype 10A-q  ---------- ---------- ---------- ---------- ----------
           Serotype 4  ---------- t--------- ---------- ---------- ----------
         Serotype 35F  ---------- ---------- ---------- ---------- ----------
           Serotype 3  ---tg----- g-----t--g --t-----a- -a--g----- ---------a
            Consensus  GAACCATTGT CTCTACCTCT CACCGTCGCA AGGGCATGTT TGAAACTCCG

                       651                                                  700
         Serotype 25F  -----c---- -t--g----- t----g-gc- --a-a-a--- ------ga--
          Serotype 38  -----c---- -t--g----- t----g-gc- --a-a-a--- ------ga--
         Serotype 19A  ---------- ---------- ---------- ---------- ----------
         Serotype 23B  ---------- ---------- ---------- ---------- ----------
  Serosubtype 6A-6B-q  ---------- ---------- ---------- ---------- ----------
```

Continue next page

Serosubtype 6B-q	----------	----------	----------	----------	----------
Serotype 11B	----------	----------	----------	----------	----------
Serotype 11A-q	----------	----------	----------	----------	----------
Serosubtype 6A-c	----------	----------	----------	----------	----------
Serosubtype 6A-ca	----------	----------	----------	----------	----------
Serosubtype 6A-g	----------	-----a----	----------	----------	----------
Serosubtype 15A-ca2	----------	----------	----------	----------	----------
Serosubtype 23F-c	----------	----------	----------	----------	----------
Serotype 18B	----------	----------	----------	----------	-t--a--a--
Serotype 18C	----------	----------	----------	----------	-t--a--a--
Serotype 19F	----------	----------	----------	----------	-t--a--a--
Serotype 18F	----------	----------	----------	---a------	-t--a--a--
Serotype 1	----------	----------	----------	----------	-t--a--a--
Serotype 18A	-----a--a-	-t---ac---	t---a--a--	-----t----	-t--a--a--
Serotype 13	-----a--a-	-t---ac---	t---a--a--	-----t----	-t--a--a--
Serotype 20	-----a--a-	-t---ac---	t---a--a--	-----t----	-t--a--a--
Serotype 9N	----------	----------	----------	----------	-t--a--a--
Serosubtype 15B-c	----------	----------	----------	----------	-t--a--a--
Serotype 16F	----------	----------	----------	----------	-t--a--a--
Serosubtype 23A-23F	----------	----------	----------	----------	-t--a--a--
Serosubtype 23F-23A	----------	----------	----------	----------	-t--a--a--
Serosubtype 15B-q	----------	----------	----------	----------	-t--a--a--
Serosubtype 15C-q	----------	----------	----------	----------	-t--a--a--
Serosubtype 15C-ca	----------	----------	----------	----------	-t--a--a--
Serosubtype 10A-23F	----------	----------	----------	----------	-t--a--a--
Serosubtype 23F-g	----------	----------	----------	----------	-t--a--a--
Serosubtype 14-g	----------	----------	----------	----------	-t--a--a--
Serotype 29	----------	----------	----------	----------	-t--a--a--
Serotype 7F	----------	----------	----------	----------	-t--a--a--
Serosubtype 14-c	----------	----------	----------	----------	-t--a--a--
Serosubtype 5-q	----------	----------	----------	----------	----------
Serosubtype 2-g	----------	----------	----------	----------	-t--a--a--
Serotype 41F	----------	----------	----------	----------	-t--a--a--
Serotype 31	----------	----------	----------	----------	-t--a--a--
Serotype 42	----------	----------	----------	----------	-t--a--a--
Serosubtype 5-c	----------	----------	----------	----------	-t--a--a--
Serotype 8	----------	---------g	----------	----------	-t--a--a--
Serotype 33B	----------	----------	----------	----------	-------a--
Serosubtype 33F-q	----------	----------	----------	----------	----------
Serosubtype 11A-nz	----------	----------	----------	----------	----------
Serosubtype 15B-22F	----------	----------	----------	----------	----------
Serotype 22F	----------	----------	----------	----------	----------
Serotype 22A	----------	----------	----------	----------	----------
Serosubtype 15A-ca1	----------	----------	----------	----------	----------
Serotype 7C	----------	----------	----------	----------	----------
Serotype 9V	----------	----------	----------	----------	----------
Serosubtype 6B-c	----------	----------	----------	----------	----------
Serotype 21	----------	----------	----------	----------	----------
Serotype 10F	----------	----------	----------	----------	----------
Serotype 12F	----------	----------	----------	----------	----------
Serosubtype 2-q	----------	----------	----------	----------	----------
Serosubtype 6A-6B-g	----------	----------	----------	----------	----------
Serosubtype 6B-g	----------	----------	----------	----------	----------
Serosubtype 23A-ca	----------	----------	----------	----------	-t--a--a--
Serotype 37	----------	----------	----------	----------	-t--a--a--
Serotype 17A	----------	----------	----------	----------	----------
Serotype 34	----------	----------	----------	----------	----------
Serosubtype 17F-35B	----------	----------	----------	----------	----------
Serotype 35B	----------	----------	----------	----------	----------
Serotype 33A	----------	----------	----------	----------	----------

Continue next page

```
Serosubtype 33F-g  ---------- ---------- ---------- ---------- ----------
Serosubtype 17F-c  ---------- ---------- ---------- ---------- ----------
Serosubtype 10A-q  ---------- --t------- ---------- ---------- ----------
       Serotype 4  ---------- ---------- ---------- ---------- ----------
     Serotype 35F  ---------- ---------- ---------- ---------- ----------
       Serotype 3  -----a--a- -t-t-ac--- t---a--a-- -----t---- -t--a--a--
        Consensus  GAAGAGAAGA TAGCAGAAAA CTTTCTTCAG GTTCGGGAAA TAGCTAAGGA

                   701                                                750
     Serotype 25F  ta-t--aga- --t---aca- --tta----- ---a-----c -----t--t-
      Serotype 38  ta-t--aga- --t---aca- --tta----- ---a-----c -----t--t-
     Serotype 19A  ---------- -----g---- -------c-- ---------- ---------t
     Serotype 23B  ---c------ -----g---- ---------- ---c------ ----------
Serosubtype 6A-6B-q  ---c------ -----g---- ---------- ---------- --------g-
  Serosubtype 6B-q  ---c------ -----g---- ---------- ---------- --------g-
     Serotype 11B  ---c------ -----g---- ---------- ---------- ----------
    Serotype 11A-q  ---c------ ---------- ---------- ---------- ----------
  Serosubtype 6A-c  ---------- ---------- -------c-- ---------- ----------
 Serosubtype 6A-ca  ---------- ---------- -------c-- ---------- ----------
  Serosubtype 6A-g  ---------- ---------- -------c-- ---------- ---------
                                                                        -
Serosubtype 15A-ca2  ---------- -----g---- -------c-- ---------- ----------
Serosubtype 23F-c  ---------- -----g---- -------c-- ---------- ----------
     Serotype 18B  ---a--aga- --t------- ---------- c--a--g--a -----t--ta
     Serotype 18C  ---a--aga- --t------- ---------- c--a--g--a -----t--ta
     Serotype 19F  ------aga- --t------- ---------- c--a--g--a -----t--t-
     Serotype 18F  ---a--aga- --t------- -c-------- c--a--g--a -----t--t-
       Serotype 1  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
     Serotype 18A  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
      Serotype 13  ---a--aga- --t------- -a-------- c--a--g--a -----t--t-
      Serotype 20  ---a--aga- --t------- -a-------- c--a--g--a -----t--t-
      Serotype 9N  ---a--aga- --t------- -a-------- c--a--g--a -----t--t-
Serosubtype 15B-c  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
     Serotype 16F  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
Serosubtype 23A-23F  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
Serosubtype 23F-23A  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
Serosubtype 15B-q  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
Serosubtype 15C-q  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
Serosubtype 15C-ca  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
Serosubtype 10A-23F  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
Serosubtype 23F-g  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
Serosubtype 14-g  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
      Serotype 29  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
      Serotype 7F  ---a--aga- --t------- -c-------- a--a--g--a -----t--t-
 Serosubtype 14-c  ---a--aga- --t------- ---------- c--a--g--a -----t--t-
  Serosubtype 5-q  ---------- -----g---- ---------- c--a--g--a -----t--t-
  Serosubtype 2-g  ---acga-a- --t------- -------c-- ---------- ---------
                                                                        -
     Serotype 41F  ---a--aga- --t------- -------c-- ---------- ----------
      Serotype 31  ---a--aga- --t------- -------c-- ---------- ----------
      Serotype 42  ---a--aga- --t------- -------c-- ---------- ----------
  Serosubtype 5-c  ---a--aga- --t------- -------c-- ---------- ----------
       Serotype 8  ---a--aga- --t------- -------c-- ---------- ----------
     Serotype 33B  ---------- -----ga--- -------c-- ---------- ----------
Serosubtype 33F-q  ---------- -----ga--- -------c-- ---------- ----------
Serosubtype 11A-nz  ---------- -----g---- -------c-- ---------- ----------
Serosubtype 15B-22F  ---------- -----g---- -------c-- a--------- ----------
     Serotype 22F  ---------- -----g---- -------c-- a--------- ----------
     Serotype 22A  ---------- -----g---- -------c-- a--------- ----------
Serosubtype 15A-ca1  ---------- -----g---- -------c-- ---------- ----------
      Serotype 7C  ---------- ---------- -------c-- a--------- ----------
      Serotype 9V  ---------- -----g---- -------c-- ---------- ----------
```

Continue next page

Serosubtype 6B-c	----------	----------	-------c--	----------	----------
Serotype 21	----------	-----ga---	-------c--	----------	----------
Serotype 10F	---c------	-----g----	----------	----------	---------t
Serotype 12F	----------	--t-------	-c-----c--	----------	----------
Serosubtype 2-q	----------	-----g----	-------c--	----------	----------
Serosubtype 6A-6B-g	----------	----------	-------c--	----------	----------
Serosubtype 6B-g	----------	----------	-------c--	----------	----------
Serosubtype 23A-ca	---a--aga-	--t-------	----------	c--a--g--a	-----t--t-
Serotype 37	---a--aga-	--t-------	----------	c--a--g--a	--t--t--t-
Serotype 17A	----------	-----g----	-------c--	----------	----------
Serotype 34	----------	-----g----	-------c--	----------	----------
Serosubtype 17F-35B	----------	-----g----	-------c--	----------	----------
Serotype 35B	----------	-----g----	-------c--	----------	----------
Serotype 33A	----------	-----g----	----------	----------	----------
Serosubtype 33F-g	----------	-----g----	----------	----------	----------
Serosubtype 17F-c	----------	-----g----	-------c--	----------	----------
Serosubtype 10A-q	----------	-----g----	-------c--	----------	----------
Serotype 4	----------	-----g----	-------c--	----------	--t-------
Serotype 35F	----------	-----g----	-------c--	----------	----------
Serotype 3	---a--aga-	--t-------	----------	c--a--g--a	-----t--t-
Consensus	AGTGGCGAGT	GACTTAGTCA	TTGCTTATGG	GGCTGAAATT	TACTACACAC

	751				800
Serotype 25F	a---ca--a-	ca----a--t	-----g--a.	actt-c--a-	-tt-a-----
Serotype 38	a---ca--a-	ca----a--t	-----g--a.	actt-c--a-	-tt-a-----
Serotype 19A	--------t-	----------	--------t-	-.--------	--------a-
Serotype 23B	--------t-	--g-------	--------t-	-.--------	--------a-
Serosubtype 6A-6B-q	--------t-	----------	-----c--t-	-.--------	--------a-
Serosubtype 6B-q	--------t-	----------	-----c--t-	-.--------	--------a-
Serotype 11B	--------t-	----------	-----t--t-	-.--------	--------a-
Serotype 11A-q	--------t-	----------	--------t-	-.--------	--------a-
Serosubtype 6A-c	----------	----------	----------	-.--------	----------
Serosubtype 6A-ca	----------	----------	----------	-.--------	----------
Serosubtype 6A-g	----------	----------	----------	-.--------	----------
Serosubtype 15A-ca2	----------	---------a	----------	-.--------	----------
Serosubtype 23F-c	----------	----------	----------	-.--------	----------
Serotype 18B	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serotype 18C	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serotype 19F	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serotype 18F	tg----c---	a--a-----a	--------ag	a.a--c--t-	----t-----
Serotype 1	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serotype 18A	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serotype 13	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serotype 20	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serotype 9N	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serosubtype 15B-c	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serotype 16F	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serosubtype 23A-23F	tg----c---	a--a-----a	--------ag	a.a-----t-	----------
Serosubtype 23F-23A	tg----c---	a--a-----a	--------ag	a.a-----t-	----------
Serosubtype 15B-q	tg----c---	a--a-----a	--------ag	a.a-----t-	----------
Serosubtype 15C-q	tg----c---	a--a-----a	--------ag	a.a-----t-	----------
Serosubtype 15C-ca	tg--------	a--a-----a	--------ag	a.a-----t-	----------
Serosubtype 10A-23F	tg----c---	a--a-----a	--------ag	a.a-----t-	----------
Serosubtype 23F-g	tg----c---	a--a-----a	--------ag	a.a-----t-	----------
Serosubtype 14-g	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serotype 29	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serotype 7F	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serosubtype 14-c	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serosubtype 5-q	tg----c---	a--a-----a	--------ag	a.a-----t-	----t-----
Serosubtype 2-g	----------	----------	----------	-.--------	----------

Continue next page

Serotype 41F	----------	----------	----------	-.--------	----------
Serotype 31	----------	----------	----------	-.--------	----------
Serotype 42	----------	----------	----------	-.--------	----------
Serosubtype 5-c	----------	----------	----------	-.--------	----------
Serotype 8	----------	----------	----------	-.--------	----------
Serotype 33B	----------	----------	----------	-.--------	----------
Serosubtype 33F-q	----------	----------	----------	-.--------	----------
Serosubtype 11A-nz	----------	----------	----------	-.--------	----------
Serosubtype 15B-22F	----------	----------	----------	-.--------	----------
Serotype 22F	----------	----------	----------	-.--------	----------
Serotype 22A	----------	----------	----------	-.--------	----------
Serosubtype 15A-ca1	----------	---------a	----------	-.--------	----------
Serotype 7C	----------	----------	----------	-.--------	----------
Serotype 9V	----------	----------	----------	-.--------	----------
Serosubtype 6B-c	----------	----------	----------	-.--------	----------
Serotype 21	----------	----------	----------	-.--------	----------
Serotype 10F	--------t-	----------	-----c--t-	-.--------	-a------a-
Serotype 12F	----------	---------a	----------	-.--------	----------
Serosubtype 2-q	----------	---------a	----------	-.--------	----------
Serosubtype 6A-6B-g	----------	----------	----------	-.--------	----------
Serosubtype 6B-g	----------	----------	----------	-.--------	----------
Serosubtype 23A-ca	tg----c---	a--a-----a	-.------aa	--a-----t-	----------
Serotype 37	tg----c---	a--a-----a	--------aa	--a-----t-	----t-----
Serotype 17A	----------	----------	----------	-.--------	----------
Serotype 34	----------	----------	----------	-.--------	----------
Serosubtype 17F-35B	----------	----------	----------	-.--------	----------
Serotype 35B	----------	----------	----------	-.--------	----------
Serotype 33A	----------	----------	----------	-.--------	----------
Serosubtype 33F-g	----------	----------	----------	-.--------	----------
Serosubtype 17F-c	----------	----------	----------	-.--------	----------
Serosubtype 10A-q	----------	----------	----------	-.--------	----------
Serotype 4	----------	----------	----------	-.--------	----------
Serotype 35F	----------	----------	----------	-.--------	----------
Serotype 3	tg----c---	a--a-----a	--------..	--a-----t-	----t-----
Consensus	CAGATGTTCT	GGATAAGCTG	GAAAAAAAGC	GAGATTCCGA	CCCTCAATGA

Figure 2

1	atgaaattga	agtttcttat	aacaaattta	tttcatgtct	ttttgtctaa	tctgattaca
61	attgtcacat	cggttatagt	tgtactaatt	ttaccaaaaa	ttatgggagt	aactgagtat
121	agttattggc	aactatatat	tttttaccta	acatatattg	gttttttca	tctgggttgg
181	attgatggaa	tttatcttaa	atatggtgga	ttagagtacc	agaatttaga	taagaaacag
241	ttttattctc	aaatacttca	atttttcagt	tttttaattt	taatttcttt	tctattattt
301	ggttttaact	tattgattgt	gacagatcca	aatgcaaaat	atatttataa	catgactatt
361	attagtatga	tagttacaaa	tttaagaatg	ttatttgttt	atattttgca	gatgacaaat
421	cgattaaagg	atagctctat	aattctgata	agtgatcgcg	ttatatatat	ttttctttta
481	tttctgttta	ttatatttaa	atggcatgaa	tacaaggtaa	tgatttgggc	ggatgtttta
541	ggaaggacat	tttctctcct	actttctttt	tggacttgta	aagatattgt	ttttcaatcc
601	ttatccgagt	tcatattgga	tctgagagag	tcttttgaca	atatccgtgt	tggaatcaac
661	ttaatgttat	ccaatattgc	aagtagtatg	attattggta	ttgttcgaat	gggaattcaa
721	tggaattgga	atatcgaaac	attcgggaaa	gtatcactga	tgctaagcat	ctctaattta
781	ttaatgactt	ttattaatgc	gattggttta	gttgtctttc	ctttgttaaa	acggacaaaa
841	acggaaaatt	tatctaaaat	ttattccaac	ttaagaaatg	ttttgatgct	gatcatgttt
901	gcaatattgc	tcttttatta	tcctttaaaa	attattctag	atctttggtt	gccagcttat
961	cgggatgcgt	tgatttttat	ggctcttatt	tttcctatgt	caatttatga	agggaagat

Figure 3

```
1    atgaaattga agtttcttat aacaaatttg tttcatgtct ttttgtctaa tctgattaca
61   attgtcacat cggttatagt tgtactaatc ttaccaaaaa ttatgggagt aactgagtat
121  agttattggc aactatatat tttttaccta acatatattg gtttttttca tctgggttgg
181  attgatggaa tttatcttaa atatggtgga ttagagtacc agaatttaga taagaaacag
241  ttttattctc aaatacttca attttccagt tttttaattt taatttcttt tctattattt
301  ggttttaact tattgattgt gacagatcca aatgcaaaat atatttataa catgactatt
361  attagtatga tagttacaaa tttaagaatg ttattcgttt atattttgca gatgacaaat
421  cgattaaagg atagctctat aattctgata agtgatcgcg ttatatatat ttttctttta
481  tttctgttta ttatatttaa atggcatgaa tacaaggtaa tgatttgggc ggatgtttta
541  ggaaggacat tttctctcct actttctttt tggatttgta aagatattgt tttcaatcc
601  ttatccgagt tcatattgga tctgagagag tcttttgaca atatccgtgt tggaatcaat
661  ttaatgttat ccaatattgc aagtagtatg attattggta ttgttcgaat gggaattcaa
721  tggaattgga atatcgaaac attcgggaaa gtatcactga cgctaagcat ctctaattta
781  ttaatgactt ttattaatgc gattggttta gttgtctttc ctttgttaaa acggacaaaa
841  acggaaaatt tatctaaaat ttattccaac ttaagaaatg ttttgatgct gatcatgttt
901  gcaatattgc tcttttatta tcctttaaaa attattctag atctttggtt gccagcttat
961  cgggatgcgt tgatttttat ggctcttatt tttcctatgt caatttatga agggaagat
```

Figure 4

1	atgaaattga	agtttcttat	aacaaatttg	tttcatgtct	ttttgtctaa	tctgattaca
61	attgtcacat	cggttatagt	tgtactaatt	ttaccaaaaa	ttatgggagt	aactgagtat
121	agttattggc	aactatatat	tttttaccta	acatatattg	gtttttttca	tctgggttgg
181	attgatggaa	tttatcttaa	atatggtgga	ttagagtacc	agaatttaga	taagaaacag
241	ttttattctc	aaatacttca	attttccagt	tttttaattt	taatttcttt	tctattattt
301	ggttttaact	tattgattgt	gacagatcca	aatgcaaaat	atatttataa	catgaccatt
361	attagtatga	tagttacaaa	tttaagaatg	ttattcgttt	atattttgca	gatgacaaat
421	cgattaaagg	atagctctat	aattctgata	agtgatcgcg	ttatatatat	ttttctttta
481	tttctgttta	ttatatttaa	atggcatgaa	tacaaggtaa	tgatttgggc	ggatgtttta
541	ggaaggacat	tttctctcct	actttctttt	tggatttgta	aagatattgt	tttcaatcc
601	ttatccgagt	tcatattgga	tctgagagag	tcttttgaca	atatccgtgt	tggaatcaat
661	ttaatgttat	ccaatattgc	aagtagtatg	attattggta	ttgttcgaat	gggaattcaa
721	tggaattgga	atatcgaaac	attcgggaaa	gtatcactga	cgctaagcat	ctctaattta
781	ttaatgactt	ttattaatgc	gattggttta	gttgtctttc	ctttgttaaa	acggacaaaa
841	acggaaaatt	tatctaaaat	ttattccaac	ttaagaaatg	ttttgatgct	gatcatgttt
901	gcaatattgc	tcttttatta	tcctttaaaa	attattctag	atctttggtt	gccagcttat
961	cgggatgcgt	tgatttttat	ggctcttatt	tttcctatgt	caatttatga	agggaagat

Figure 5

1	atgaaattga	agtttcttat	aacaaatttg	tttcatgtct	ttttgtctaa	tctgattaca
61	attgtcacat	cggttatagt	tgtactaatc	ttaccaaaaa	ttatgggagt	aactgagtat
121	agttattggc	aactatatat	tttttaccta	acatatattg	gtttttttca	tctgggttgg
181	attgatggaa	tttatcttaa	atatggtgga	ttagagtacc	agaatttaga	taagaaacag
241	ttttattctc	aaatacttca	atttttcagt	tttttaattt	taatttcttt	tctattattt
301	ggttttaact	tattgattgt	gacagatcca	aatgcaaaat	atatttataa	catgactatt
361	attagtatga	tagttacaaa	tttaagaatg	ttattcgttt	atattttgca	gatgacaaat
421	cgattaaagg	atagctctat	aattctgata	agtgatcgcg	ttatatatat	ttttctttta
481	tttctgttta	ttatatttaa	atggcatgaa	tacaaggtaa	tgatttgggc	ggatgtttta
541	ggaaggacat	tttctctcct	actttctttt	tggatttgta	aagatattgt	ttttcaatcc
601	ttatccgagt	tcatattgga	tctgagagag	tcttttgaca	atatccgtgt	tggaatcaat
661	ttaatgttat	ccaatattgc	aagtagtatg	attattggta	ttgttcgaat	gggaattcaa
721	tggaattgga	atatcgaaac	attcgggaaa	gtatcactga	cgctaagcat	ctctaattta
781	ttaatgactt	ttattaatgc	gattggttta	gttgtctttc	ctttgttaaa	acggacaaaa
841	acggaaaatt	tatctaaaat	ttattccaac	ttaagaaatg	ttttgatgct	gatcatgttt
901	gcaatattgc	tcttttatta	tcctttaaaa	attattctag	atctttggtt	gccagcttat
961	cgggatgcgt	tgatttttat	ggctcttatt	tttcctatgt	caatttatga	agggaagat

Figure 6

```
1     atgaaattga agtttcttat aacaaattta tttcatgtct ttttgtctaa tctgattaca
61    attgtcacat cggttatagt tgtactaatt ttaccaaaaa ttatgggagt aactgagtat
121   agttattggc aactatatat tttttaccta acatatattg gtttttttca tctgggttgg
181   attgatggaa tttatcttaa atatggtgga ttagagtacc agaatttaga taagaaacag
241   ttttattctc aaatacttca atttttcagt tttttaattt taatttcttt tctattattt
301   ggttttaact tattgattgt gacagatcca aatgcaaaat atatttataa catgactatt
361   attagtatga tagttacaaa tttaagaatg ttatttgttt atattttgca gatgacaaat
421   cgattaaagg atagctctat aattctgata agtgatcgcg ttatatatat ttttctttta
481   tttctgttta ttatatttaa atggcatgaa tacaaggtaa tgatttgggc ggatgtttta
541   ggaaggacat tttctctcct actttctttt tggacttgta aagatattgt ttttcaatcc
601   ttatccgagt tcatattgga tctgagagag tcttttgaca atatccgtgt tggaatcaac
661   ttaatgttat ccaatattgc aagtagtatg attattggta ttgttcgaat gggaattcaa
721   tggaattgga atatcgaaac attcgggaaa gtatcactga tgctaagcat ctctaattta
781   ttaatgactt ttattaatgc gattggttta gttgtctttc ctttgttaaa acggacaaaa
841   acggaaaatt tatctaaaat ttattccaac ttaagaaatg ttttgatgct gatcatgttt
901   gcaatattgc tcttttatta tcctttaaaa attattctag atctttggtt gccagcttat
961   cgggatgcgt tgatttttat ggctcttatt tttcctatgt caatttatga agggaagat
```

Figure 7

1	atgaaattga	agtttcttat	aacaaatttg	tttcatgtct	ttttgtctaa	tctgattaca
61	attgtcacat	cggttatagt	tgtactaatc	ttaccaaaaa	ttatgggagt	aactgagtat
121	agttattggc	aactatatat	tttttaccta	acatatattg	gtttttttca	tctgggttgg
181	attgatggaa	tttatcttaa	atatggtgga	ttagagtacc	agaatttaga	taagaaacag
241	ttttattctc	aaatacttca	atttttcagt	tttttaattt	taatttcttt	tctattattt
301	ggttttaact	tattgattgt	gacagatcca	aatgcaaaat	atatttataa	catgactatt
361	attagtatga	tagttacaaa	tttaagaatg	ttattcgttt	atattttgca	gatgacaaat
421	cgattaaagg	atagctctat	aattctgata	agtgatcgcg	ttatatatat	ttttctttta
481	tttctgttta	ttatatttaa	atggcatgaa	tacaaggtaa	tgatttgggc	ggatgtttta
541	ggaaggacat	tttctctcct	actttctttt	tggatttgta	aagatattgt	ttttcaatcc
601	ttatccgagt	tcatattgga	tctgagagag	tcttttgaca	atatccgtgt	tggaatcaat
661	ttaatgttat	ccaatattgc	aagtagtatg	attattggta	ttgttcgaat	gggaattcaa
721	tggaattgga	atatcgaaac	attcgggaaa	gtatcactga	cgctaaacat	ctctaattta
781	ttaatgactt	ttattaatgc	gattggttta	gttgtctttc	ctttgttaaa	acggacaaaa
841	acggaaaatt	tatctaaaat	ttattccaac	ttaagaaatg	ttttgatgct	gatcatgttt
901	gcaatattgc	tcttttatta	tcctttaaaa	attattctag	atctttggtt	gccagcttat
961	cgggatgcgt	tgatttttat	ggctcttatt	tttcctatgt	caatttatga	agggaagat

Figure 8

```
1     atgaaattga agtttcttat aacaaatttg tttcatgttc ttttgtctaa tctgattaca
61    attcttacat cagttatagt tgtactaatt ttaccaaaaa ttatgggagt aactgagtat
121   agttattggc aactatatat tttttaccta acatatattg gtttttttca tctgggatgg
181   attgatggaa tttatcttaa atatggcgga ttagagtacc agaacttaga taagaaacag
241   ttttattctc aaatacttca attttccagt tttttaattt taatttcttt tctattattt
301   ggttttaact tattgactgt gacagatcaa aatgcaaaat atatttataa catgactatt
361   attagtatga tagttacaaa tttaagaatg ttattcgttt atattttgca gatgacaaat
421   cgattaaagg atagttccat cattctaatc agtgatcgcg ttatatatgt tattctttta
481   ttcctgttta ttatatttaa atggcatgaa tacaaggtaa tgatttgggc agatgttttg
541   ggaaggacat tttctctcct actttctttt tggatttgta aagatattgt tttttcaatcc
601   ttatccgagt ttatattgga tctgagagag tcttttgaca atatccgtgt tggaatcaat
661   ttaatgttat ccaatattgc aagtagtatg attattggta ttgttcgaat gggaattcaa
721   tggaattgga atatcgaaac attcgggaaa gtatcactga cgctaagcat ctctaattta
781   ttaatgactt ttattaatgc gattggttta gttgtttttc ctttgttaaa acggacaaaa
841   acggaaaatt tatctaaaat ttattccaac ttaagaaatg ttttgatgct tatcatgttc
901   gcgattttgc tcatttacta tcctttaaaa attgtattag acctctggtt gccagcctat
961   caagatgcct tgattttcat ggctcttatt tttcctatgt caatttatga agggaagat
```

Figure 9

1	atgaaattga	agtttcttat	aacaaatttg	tttcatgtct	ttttatctaa	tctgattaca
61	attgtcacat	cggttatagt	tgtactaatt	ttaccaaaaa	ttatgggagt	aactgagtat
121	agttattggc	aactatatat	tttttaccta	acatatattg	gttttttca	tctgggttgg
181	attgatggaa	tttatcttaa	atatggtgga	ttagagtacc	agaatttaga	taagaaacag
241	ttttattctc	aaatacttca	atttttcagt	tttttaattt	taatttcttt	tctattattt
301	ggttttaact	tattgattgt	gacagatcca	aatgcaaaat	atatttataa	catgactatt
361	attagtatga	tagttacaaa	tttaagaatg	ttattcgttt	atattttgca	gatgacaaat
421	cgattaaagg	atagctctat	aattctgata	agtgatcgcg	tcatatatat	ttttctttta
481	tttctgttta	ttatatttaa	atggcatgaa	tacaaggtaa	tgatttgggc	ggatgtttta
541	ggaaggacat	tttctctcct	actttctttt	tggatttgta	aagatattgt	tttttcaatcc
601	ttatccgagt	tcatattgga	tctgagagaa	tcttttgaca	atatccgtgt	tggaatcaat
661	ttaatgttat	ccaatattgc	aagtagtatg	attattggta	ttgttcgaat	gggaattcaa
721	tggaattgga	atatcgaaac	attcgggaaa	gtatcactga	cgctaagcat	ctctaattta
781	ttaatgactt	ttattaatgc	gattggttta	gttgtctttc	ctttgttaaa	acggacaaaa
841	acggaaaatt	tatctaaaat	ttattccaac	ttaagaaatg	ttttgatgct	gatcatgttt
901	gcaatattac	tcttttatta	tcctttaaaa	attattctag	atctttggtt	gccagcttat
961	cgggatgcgt	tgatttttat	ggctcttatt	tttcctatgt	caatttatga	agggaagat

Figure 10

```
1     atgaaattga agtttcttat aacaaatttg tttcatgtct ttttatctaa tctgattaca
61    attgtcacat cggttatagt tgtactaatt ttaccaaaaa ttatgggagt aactgagtat
121   agttattggc aactatatat tttttaccta acatatattg gtttttttca tctgggttgg
181   attgatggaa tttatcttaa atatggtgga ttagagtacc agaatttaga taagaaacag
241   ttttattctc aaatacttca atttttcagt tttttaattt taatttcttt tctattattt
301   ggttttaact tattgattgt gacagatcca aatgcgaaat atatttataa catgactatt
361   attagtatga tagttacaaa tttaagaatg ttattcgttt atattttgca gatgacaaat
421   cgattaaagg atagctctat aattctgata agtgatcgcg tcatatatat ttttctttta
481   tttctgttta ttatatttaa atggcatgaa tacaaggtaa tgatttgggc ggatgtttta
541   ggaaggacat tttctctcct actttctttt tggatttgta aagatattgt ttttcaatcc
601   ttatccgagt tcatattgga tctgagagaa tcttttgaca atatccgtgt tggaatcaat
661   ttaatgttat ccaatattgc aagtagtatg attattggta ttgttcgaat gggaattcaa
721   tggaattgga atatcgaaac attcgggaaa gtatcactga cgctaagcat ctctaattta
781   ttaatgactt ttattaatgc gattggttta gttgtctttc ctttgttaaa acggacaaaa
841   acggaaaatt tatctaaaat ttattccaac ttaagaaatg ttttgatgct gatcatgttt
901   gcaatattac tcttttatta tcctttaaaa attattctag atctttggtt gccagcttat
961   cgggatgcgt tgatttttat ggctcttatt tttcctatgt caatttatga agggaagat
```

Figure 11

```
1     atgaaattga agtttcttat aacaaatttg tttcatgttc ttttgtctaa tctgattaca
61    attcttacat cagttatagt tgtactaatt ttaccaaaaa ttatgggagt aactgagtat
121   agttattggc aactatatat tttttaccta acatatattg gtttttttca tctgggatgg
181   attgatggaa tttatcttaa atatggcgga ttagagtacc agaacttaga taagaaacag
241   ttttattctc aaatacttca attttccagt tttttaattt taatttcttt tctattattt
301   ggttttaact tattgactgt gacagatcaa aatgcaaaat atatttataa catgactatt
361   attagtatga tagttacaaa tttaagaatg ttattcgttt atattttgca gatgacaaat
421   cgattaaagg atagttccat cattctaatc agtgatcgcg ttatatatgt tattctttta
481   ttcctgttta ttatatttaa atggcatgaa tacaaggtaa tgatttgggc agatgttttg
541   ggaaggacat tttctctcct actttctttt tggatttgta aagatattgt ttttcaatcc
601   ttatccgagt ttatattgga tctgagagag tcttttgaca atatccgtgt tggaatcaat
661   ttaatgttat ccaatattgc aagtagtatg attattggta ttgttcgaat gggaattcaa
721   tggaattgga atatcgaaac attcgggaaa gtatcactga cgctaagcat ctctaattta
781   ttaatgactt ttattaatgc gattggttta gttgtttttc ctttgttaaa acggacaaaa
841   acggaaaatt tatctaaaat ttattccaac ttaagaaatg ttttgatgct tatcatgttc
901   gcgattttgc tcatttacta tcctttaaaa attgtattag acctctggtt gccagcctat
961   caagatgcct tgattttcat ggctcttatt tttcctatgt caatttatga agggaagat
```

Figure 12

```
1     atgcttttaa atttcttatt catatctatt tttctattaa ttatcattac atttatatta
61    tttgaggggg atttttttca acctgcagta attttaacac tcacttattt tatttcgatt
121   gcaagtgctc tagttaatag aaatgtttgg ggaacagaac tccatttcaa aacctttggt
181   ttgatattgt taggggttgc tacatttatt atagtttcct tgttgacaaa attgtcgtac
241   aggcctaaag tggagggaat ttcgtatgaa gaattgaaag aaataaatcc ttcaaagata
301   atctatgtca ttcttctgat tctaaatctt gttatgctat ttctttatac ccgtgaaatt
361   cagaaagtgg tattgttttc aggtagaagt ttttctaata ttacagattt gataagtaac
421   tataggtacc tatcttatta ttcaaatgaa gtagaaataa gtggaatgat taatcaacta
481   tctaaaatta ttccagcgac tacacttatt tctttatata tatttataaa taattatttt
541   ataactaaac aaataaagaa aaatttcatt tatttgattc caatagctat attctttgtc
601   tatgcaatca ttagtggtgg tagattgccc cttataaggt tagttgttgg agctctgttg
661   atattgtata tatactctgt gtacgggagt cctaaatctc aacttaccaa aagttttaaa
721   atgatcactc gctctctgtt tac
```

Figure 13

```
1     atgcttttaa atttcttatt catatctatt tttctattaa ttatcattac atttatatta
61    tttgaggggg atttttttca acctgcagta attttaacaa tcgcttattt tatttcgatt
121   gcaagtgctc tagttaatag aaatgtttgg ggaacagaac tccatttcaa aacctttat
181   ttgatattgt taggggttgc tacatttgtt atagtttcct tgttgacaaa attgtcgtac
241   aggcctaaag tggagggaat ttcgcatgaa gaattgaaag aaataaatcc ttcaaagata
301   atctatgtca ttcttctgac tctaaatctt gttatgttat ttctttatat ccgtgaaatt
361   cagaaagtag tattgttttc aggtagaagt ttttctaata ttacagattt gataagtaac
421   tataggtacc tatcttatta ttcaaatgaa gtagaaaatc gtgtaagtgg aatgattaat
481   caactatcta aaattattcc agcgactaca cttatttctt tatatatatt tatgaataat
541   tattttataa ctaaacaaat aaagaaaaat ttcatttatt tgattccaat agctatattc
601   tttgtctatg caatcattag tggtggtaga ttgcccctta taaggttagt tgttggatct
661   ctgttgatat tgtatatata ctctgtgtac gggagtccta aatctcaact taccaaaagt
721   tttaaaatga tcactcgctc tctgtttac
```

Figure 14

```
1     atgcttttaa atttcttatt catatctatt tttctattaa ttatcattac atttatatta
61    tttgaggggg attttttca acctgcagta attttaacaa tcgcttattt tatttcgatt
121   gcaagtgctc tagttaatag aaatgtttgg ggaacagaac tccatttcaa aacctttat
181   ttgatattgt taggggttgc tacatttatt atagtttcct tgttgacaaa attgtcgtac
241   aggcctaaag tggagggaat ttcgcatgaa gaattgaaag aaataaatcc ttcaaagata
301   atctatgtca ttcttctgat tctaaatctt gttatgctat ttctttatat ccgtgaaatt
361   cagaaagtgg tattgttttc aggtagaagt ttttctaata ttacagattt gataagtaac
421   tataggtacc tatcttatta ttcaaatgaa gtagaaaatc gtgtaagtgg aatgattaat
481   caactatcta aaattattcc agcgactaca cttatttctt tatatatatt tataaataat
541   tattttataa ctaaacaaat aaagaaaaac ttcatttatt tgattccaat agctatattc
601   tttgtctatg caatcattag tggtggtaga ttgcccctta taaggttagt tgttggagct
661   ctgttgatat tgtatatata ctctgtgtac gggagtccta aatctcaact taccaaaagt
721   tttaaaatga tcactcgctc tctgtttac
```

Figure 15

1	atgcttttaa	atttcttatt	catatctatt	tttctattaa	ttatcattac	atttatatta
61	tttgaggggg	attttttca	acctgcagta	attttaacaa	tcgcttattt	tatttcgatt
121	gcaagtgctc	tagttaatag	aaatgtttgg	ggaacagaac	tccatttcaa	aacctttat
181	ttgatattgt	taggggttgc	tacatttgtt	atagtttcct	tgttgacaaa	attgtcgtac
241	aggcctaaag	tggagggaat	ttcgcatgaa	gaattgaaag	aaataaatcc	ttcaaagata
301	atctatgtca	ttcttctgac	tctaaatctt	gttatgttat	ttctttatat	ccgtgaaatt
361	cagaaagtag	tattgttttc	aggtagaagt	ttttctaata	ttacagattt	gataagtaac
421	tataggtacc	tatcttatta	ttcaaatgaa	gtagaaaatc	gtgtaagtgg	aatgattaat
481	caactatcta	aaattattcc	agcgactaca	cttatttctt	tatatatatt	tatgaataat
541	tattttataa	ctaaacaaat	aaagaaaaat	ttcatttatt	tgattccaat	agctatattc
601	tttgtctatg	caatcattag	tggtggtaga	ttgcccctta	taaggttagt	tgttggatct
661	ctgttgatat	tgtatatata	ctctgtgtac	gggagtccta	aatctcaact	taccaaaagt
721	tttaaaatga	tcactcgctc	tctgtttac			

Figure 16

```
1     atgcttttaa atttcttatt catatctatt tttctattaa ttatcattac atttatatta
61    tttgaggggg atttttttca acctgcagta attttaacac tcacttattt tatttcgatt
121   gcaagtgctc tagttaatag aaatgtttgg ggaacagaac tccatttcaa aacctttggt
181   ttgatattgt taggggttgc tacatttatt atagtttcct tgttgacaaa attgtcgtac
241   aggcctaaag tggagggaat ttcgtatgaa gaattgaaag aaataaatcc ttcaaagata
301   atctatgtca ttcttctgat tctaaatctt gttatgctat ttctttatac ccgtgaaatt
361   cagaaagtgg tattgttttc aggtagaagt ttttctaata ttacagattt gataagtaac
421   tataggtacc tatcttatta ttcaaatgaa gtagaaataa gtggaatgat taatcaacta
481   tctaaaatta ttccagcgac tacacttatt tctttatata tatttataaa taattatttt
541   ataactaaac aaataaagaa aaatttcatt tatttgattc caatagctat attctttgtc
601   tatgcaatca ttagtggtgg tagattgccc cttataaggt tagttgttgg agctctgttg
661   atattgtata tatactctgt gtacgggagt cctaaatctc aacttaccaa aagttttaaa
721   atgatcactc gctctctgtt tac
```

Figure 17

1	atgcttttaa	atttcttatt	catatctatt	tttctattaa	ttatcattac	atttatatta
61	tttgaggggg	atttttttca	acctgcagta	attttaacaa	tcgcttattt	tatttcgatt
121	gcaagtgctc	tagttaatag	aaatgtttgg	ggaacagaac	tccatttcaa	aaccttttat
181	ttgatattgt	taggggttgc	tacatttgtt	atagtttcct	tgttgacaaa	attgtcgtac
241	aggcctaaag	tggagggaat	ttcgcatgaa	gaattgaaag	aaataaatcc	ttcaaagata
301	atctatgtca	ttcttctgac	tctaaatctt	gttatgttat	ttctttatat	ccgtgaaatt
361	cagaaagtag	tattgttttc	aggtagaagt	ttttctaata	ttacagattt	gataagtaac
421	tataggtacc	tatcttatta	ttcaaatgaa	gtagaaaatc	gtgtaagtgg	aatgattaat
481	caactatcta	aaattattcc	agcgactaca	cttatttctt	tatatatatt	tatgaataat
541	tattttataa	ctaaacaaat	aaagaaaaat	ttcatttatt	tgattccaat	agctatattc
601	tttgtctatg	caatcattag	tggtggtaga	ttgcccctta	taaggttagt	tgttggatct
661	ctgttgatat	tgtatatata	ctctgtgtac	gggagtccta	aatctcaact	taccaaaagt
721	tttaaaatga	tcactcgctc	tctgtttac			

Figure 18

```
1     atgcttttaa atttcttatt catatctatt tttctattaa ttatcattac atttatatta
61    tttgagggag atttgtttca acccgcagta attttaacac ttgcttattt tatttcgatt
121   gcaagtgctc tagttaatag aaatgtttgg ggaacagaac tccatttcaa aacctttggt
181   ttgatattgc taggggttgc tacatttatt atagtttcct tgttgacaaa attgtcgtac
241   aaacctaaag tggagggaat ttcgtataaa gaattaaaag aaataaatcc ttcaaagata
301   atatatggca ttcttctgat tctaaatctt gttatgctat ttctttatat ccatgaaatt
361   cagaaagtgg tactgttttc aggtagaggt tttctaata ttacagattt gataagtaac
421   tataggtacc tatcttatta ttcaaatgaa gtagaagatc gtgtaagtgg aatgattaat
481   caactagcta aaattattcc agcgactaca tttgtttctt tatatatatt tataaataat
541   tattttataa cgaagcaaat aaagaaaaat ttcatttatt tgattccaat agctatattc
601   tttgtctatg caatcattag tggtggtaga ctgcccctta taaggttagt tattggaact
661   ctgttgatat tgtatatata ctctgtgtac gggagtcata aatctcaact taccaaaagt
721   tttaaaatga tcactcgctc tctgtttac
```

Figure 19

```
1     atgcttttaa atttcttatt catatctatt tttctattaa ttattattac atttatatta
61    tttgaggggg atttttttca acctgcagta attttaacaa tcgcttattt tatttcgatt
121   gcaagtgctc tagttaatag aaatgtttgg ggaacagaac tccatttcaa aacctttttat
181   ttgatattgt taggggttgt tacatttgtt atagtttcct tgttgacaaa attgtcgtac
241   aggcctaaag tggagggaat ttcgcatgaa gaattgaaag aaataaatcc ttcaaagata
301   atctatgtca ttcttctgac tctaaatctt gttatgttat ttctttatat ccgtgaaatt
361   cagaaagtag tattgttttc aggtagaagt ttttctaata ttacagattt gataagtaac
421   tataggtacc tatcttatta ttcaaatgaa gtagaaaatc gtgtaagtgg aatgattaat
481   caactatcta aaattattcc agcgactaca cttatttctt tatatatatt tatgaataat
541   tattttataa ctaaacaaat aaagaaaaat ttcatttatt tgattccaat agctatattc
601   tttgtctatg caatcattag tggtggtaga ttgcccctta taaggttagt tgttggagct
661   ctgttgatat tgtatatata ctctgtgtac gggagtccta aatctcaact taccaaaagt
721   tttaaaatga tcactcgctc tctgtttac
```

Figure 20

1	atgcttttaa	atttcttatt	catatctatt	tttctattaa	ttattattac	atttatatta
61	tttgaggggg	atttttttca	acctgcagta	attttaacaa	tcgcttattt	tatttcgatt
121	gcaagtgctc	tagttaatag	aaatgtttgg	ggaacagaac	tccatttcaa	aacctttat
181	ttgatattgt	taggggttgt	tacatttgtt	atagtttcct	tgttgacaaa	attgtcgtac
241	aggcctaaag	tggagggaat	ttcgcatgaa	gaattgaaag	aaataaatcc	ttcaaagata
301	atctatgtca	ttcttctgac	tctaaatctt	gttatgttat	ttctttatat	ccgtgaaatt
361	cagaaagtag	tattgttttc	aggtagaagt	ttttctaata	ttacagattt	gataagtaac
421	tataggtacc	tatcttatta	ttcaaatgaa	gtagaaaatc	gtgtaagtgg	aatgattaat
481	caactatcta	aaattattcc	agcgactaca	cttatttctt	tatatatatt	tatgaataat
541	tattttataa	ctaaacaaat	aaagaaaaat	ttcatttatt	tgattccaat	agctatattc
601	tttgtctatg	caatcattag	tggtggtaga	ttgcccctta	taaggttagt	tgttggagct
661	ctgttgatat	tgtatatata	ctctgtgtac	gggagtccta	aatctcaact	taccaaaagt
721	tttaaaatga	tcactcgctc	tctgtttac			

Figure 21

```
1     atgcttttaa atttcttatt catatctatt tttctattaa ttatcattac atttatatta
61    tttgagggag atttgtttca acccgcagta attttaacac ttgcttattt tatttcgatt
121   gcaagtgctc tagttaatag aaatgtttgg ggaacagaac tccatttcaa aacctttggt
181   ttgatattgc taggggttgc tacatttatt atagtttcct tgttgacaaa attgtcgtac
241   aaacctaaag tggagggaat ttcgtataaa gaattaaaag aaataaatcc ttcaaagata
301   atatatggca ttcttctgat tctaaatctt gttatgctat ttctttatat ccatgaaatt
361   cagaaagtgg tactgttttc aggtagaggt ttttctaata ttacagattt gataagtaac
421   tataggtacc tatcttatta ttcaaatgaa gtagaagatc gtgtaagtgg aatgattaat
481   caactagcta aaattattcc agcgactaca tttgtttctt tatatatatt tataaataat
541   tattttataa cgaagcaaat aaagaaaaat ttcatttatt tgattccaat agctatattc
601   tttgtctatg caatcattag tggtggtaga ctgcccctta taaggttagt tattggaact
661   ctgttgatat tgtatatata ctctgtgtac gggagtcata aatctcaact taccaaaagt
721   tttaaaatga tcactcgctc tctgtttac
```

Figure 22

```
1     tttgaaatgg ttgtggagtt atagattctt tttatttagg ttaaatggta ttaaagaagg
61    aaatagtgaa ttcgagaaga gttttaagat aattaggaga tactataaaa caggacgata
121   gaatgagtaa atataaggaa ttagcaaaaa atacaggtat ttttgctttg gctaactttt
181   catcaaagat tttaattttt ttgttagtac ctatatatac acgggtactt accactacgg
241   aatatggttt ttatgactta gtctatacaa ctattcagct ttttgtacca atcttgacat
301   taaatatatc tgaagccgtt atgaggttcc taatgaaaga tggtgtttct aaaaaatcag
361   tcttttcaat tgctgtttta gatatattta ttggatcaat tgcttttgct ttattgttgt
421   tagtaaataa cctgttttct ttatcagatt taatttctca atacagtatt tacatatttg
481   taatctttgt tttctatacc ctaaataatt ttttgataca attttctaag ggaattgata
541   aaattggtgt tacagctatc tctggggtca taagtacagc agttatgctt gccatgaatg
601   tcattcttct agtagtattt gattg
```

Figure 23

```
1     tttgaaatgg ttgtggagtt atagattctt tttatttagg ttaaatggta ttaaagaagg
61    aaatagtgaa ttcgagaaga gttttaagat aattaggaga tactataaaa caggacgata
121   gaatgagtaa atataaggaa ttagcaaaaa atacaggtat ttttgctttg gctaactttt
181   catcaaagat tttaattttt ttgttagtac ctatatatac acgggtactt accactacgg
241   aatatggttt ttatgactta gtctatacaa ctattcagct ttttgtacca atcttgacat
301   taaatatatc tgaagccgtt atgaggttcc taatgaaaga tggtgtttct aaaaaatcag
361   tcttttcaat tgctgtttta gatatattta ttggatcaat tgcttttgct ttattgttgt
421   tagtaaataa cctgttttct ttatcagatt taatttctca atacagtatt tacatatttg
481   taatctttgt tttctatacc ctaaataatt ttttgataca attttctaag ggaattgata
541   aaattggtgt tacagctatc tctggggtca taagtacagc agttatgctt gccatgaatg
601   tcattcttct agtagtattt gattg
```

Figure 24

```
1     tttgaaatgg ttgtggagtt atagattctt tttatttagg ttaaatggta ttaaagaagg
61    aaatagtgaa ttcgagaaga gttttangat aattaggaga tactataaaa caggacgata
121   gaatgagtaa atataaggaa ttagcaaaaa atacaggtat ttttgctttg gctaactttt
181   catcaaagat tttaattttt ttgttagtac ctatatatac acgggtactt accactacgg
241   aatatggttt ttatgactta gtctatacaa ctattcagct ttttgtacca atcttgacat
301   taaatatatc tgaagccgtt atgaggttcc taatgaaaga tggtgtttct aaaaaatcag
361   tcttttcaat tgctgtttta gatatattta ttggatcaat tgcttttgct ttattgttgt
421   tagtaaataa cctgttttct ttatcagatt taatttctca atacagtatt tacatatttg
481   taatctttgt tttctatacc ctaaataatt ttttgataca attttctaag ggaattgata
541   aaattggtgt tacagctatc tctgggatca taagtacagc agntatgctt gccatgaatg
601   tcattcttct agtagtattt gattg
```

Figure 25

1	tttgaaatgg	ttgtggagtt	atagattctt	tttatttagg	ttaaatggta	ttaaagaagg
61	aaatagtgaa	ttcgagaaga	gttttaagat	aattaggaga	tactataaaa	caggacgata
121	gaatgagtaa	atataaggaa	ttagcaaaaa	atacaggtat	ttttgctttg	gctaactttt
181	catcaaagat	tttaattttt	ttgttagtac	ctatatatac	acgggtactt	accactacgg
241	aatatggttt	ttatgactta	gtctatacaa	ctattcagct	ttttgtacca	atcttgacat
301	taaatatatc	tgaagccgtt	atgaggttcc	taatgaaaga	tggtgtttct	aaaaaatcag
361	tcttttcaat	tgctgtttta	gatatattta	ttggatcaat	tgcttttgct	ttattgttgt
421	tagtaaataa	cctgttttct	ttatcagatt	taatttctca	atacagtatt	tacatatttg
481	taatctttgt	tttctatacc	ctaaataatt	ttttgataca	attttctaag	ggaattgata
541	aaattggtgt	tacagctatc	tctggggtca	taagtacagc	agttatgctt	gccatgaatg
601	tcattcttct	agtagtattt	gattg			

Figure 26

```
1     tttgaaatgg ttgtggagtt atagattctt tttatttagg ttaaatggta ttaaagaagg
61    aaatagtgaa ttcgagaaga gttttaagat aattaggaga tactataaaa caggacgata
121   gaatgagtaa atataaggaa ttagcaaaaa atacaggtat ttttgctttg gctaactttt
181   catcaaagat tttaattttt ttgttagtac ctatatatac acgggtactt accactacgg
241   aatatggttt ttatgactta gtctatacaa ctattcagct ttttgtacca atcttgacat
301   taaatatatc tgaagccgtt atgaggttcc taatgaaaga tggtgtttct aaaaaatcag
361   tcttttcaat tgctgtttta gatatattta ttggatcaat tgcttttgct ttattgttgt
421   tagtaaataa cctgttttct ttatcagatt taatttctca atacagtatt tacatatttg
481   taatctttgt tttctatacc ctaaataatt ttttgataca attttctaag ggaattgata
541   aaattggtgt tacagctatc tctggggtca taagtacagc agttatgctt gccatgaatg
601   tcattcttct agtagtattt gattg
```

Figure 27

```
1     tttgaaatgg ttgtggagtt atagattctt tttatttagg ttaaatggta ttaaagaagg
61    aaatagtgaa ttcgagaaga gttttangat aattaggaga tactataaaa caggacgata
121   gaatgagtaa atataaggaa ttagcaaaaa atacaggtat ttttgctttg gctaactttt
181   catcaaagat tttaattttt ttgttagtac ctatatatac acgggtactt accactacgg
241   aatatggttt ttatgactta gtctatacaa ctattcagct ttttgtacca atcttgacat
301   taaatatatc tgaagccgtt atgaggttcc taatgaaaga tggtgtttct aaaaaatcag
361   tcttttcaat tgctgtttta gatatattta ttggatcaat tgcttttgct ttattgttgt
421   tagtaaataa cctgttttct ttatcagatt taatttctca atacagtatt tacatatttg
481   taatctttgt tttctatacc ctaaataatt ttttgataca attttctaag ggaattgata
541   aaattggtgt tacagctatc tctggggtca taagtacagc agttatgctt gccatgaatg
601   tcattcttct agtagtattt gattg
```

Figure 28

```
1     tttgaaatgg ttgtggagtt atagattctt tttatttagg ttaaatggta ttaaagaagg
61    aaatagtgaa ttcgagaaga gttttaagat aattaggaga tactataaaa caggacgata
121   gaatgagtaa atataaggaa ttagcaaaaa atacaggtat ttttgctttg gctaactttt
181   catcaaagat tttaattttt ttgttagtac ctatatatac acgggtactt accactacgg
241   aatatggttt ttatgactta gtctatacaa ctattcagct ttttgtacca atcttgacat
301   taaatatatc tgaagccgtt atgaggttcc taatgaaaga tggtgtttct aaaaaatcag
361   tcttttcaat tgctgtttta gatatattta ttggatcaat tgcttttgct ttattgttgt
421   tagtaaataa cctgttttct ttatcagatt taatttctca atacagtatt tacatatttg
481   taatctttgt tttctatacc ctaaataatt ttttgataca attttctaag ggaattgata
541   aaattggtgt tacagctatc tctggggtca taagtacagc agttatgctt gccatgaatg
601   tcattcttct agnagtattt gattg
```

Figure 29

```
1     tgcgctatta ataattttcc ttatgataac tactatattt atagagacct atatgttttt
61    atttgtcatt tctttatact attctcttga ttttggggac gatagagatt gtcatgagaa
121   acagtacatt actaattaat aataaaggtg tgaacagaaa taagaagaaa tgaaaatact
181   aaaaaactat gcctacaatc tttcttatca attgttggtg atcatactcc ctatcattac
241   gactccctat gtaacgaggg tttttagttc tgacgattta ggaacgtatg gctactttag
301   ctccattgtt acctatttta ccttgcttgc aactcttggt gttgccaact acggtaccaa
361   agagatttca gcacatcgta aggaaattgg gaagaatttc tggggaattt attctctcca
421   gtttggtgca acttggctat ccattttgct ttatcttgcc ctttgtttct tatttacttc
481   aatgcaaaat ccggtagctt atatattggg attaagttta gtgtcaaaag gtttggatat
541   ttcttggtta tttcaaggtt tggaggattt tagaaagatt acagttcgga acatcactgt
601   taagtta
```

Figure 30

```
1     tgcgctatta ataattttcc ttatgataac tactatattt atagagacct atatgttttt
61    atttgtcatt tctttatact attctcttga ttttggggac gatagagatt gtcatgagaa
121   acagtacatt actaattaat aataaaggtg tgaacagaaa taagaagaaa tgaaaatact
181   aaaaaattat gcctacaatc tttcttatca attgttggtg atcatactcc ctatcattac
241   gactccctat gtaacgaggg tttttagttc tgacgattta ggaacgtatg gctactttag
301   ctccattgtt acctatttta ccttgcttgc aactcttggt gttgccaact acggtaccaa
361   agagatttca gcacatcgta aggaaattgg gaagaatttc tggggaattt attctctcca
421   gtttggtgca acttggctat ccattttgct ttatcttgcc ctttgtttct tatttacttc
481   aatgcaaaat ccggtagctt atatattggg attaagttta gtgtcaaaag gtttggatat
541   ttcttggtta tttcaaggtt tggaggattt tagaaagatt acagttcgga acatcactgt
601   taagtta
```

Figure 31

```
1     tgcgctatta ataattttcc ttatgataac tactatattt atagagacct atatgttttt
61    atttgtcatt tctttatact attctcttga ttttggggac gatagagatt gtcatgagaa
121   acagtacatt actaattaat aataaaggtg tgaacagaaa taagaagaaa tgaaaatact
181   aaaaaattat gcctacaatc tttcttatca attgttggtg atcatactcc ctatcattac
241   gactccctat gtaacgaggg tttttagttc tgacgattta ggaacgtatg gctactttag
301   ctccattgtt acctatttta ccttgcttgc aactcttggt gttgccaact acggtaccaa
361   agagatttca gcacatcgta aggaaattgg gaagaatttc tggggaattt attctctcca
421   gtttggtgca acttggctat ccattttgct ttatcttgcc ctttgtttct tatttacttc
481   aatgcaaaat ccggtagctt atatattggg attaagttta gtgtcaaaag gtttggatat
541   ttcttggtta tttcaaggtt tggaggattt tagaaagatt acagttcgga acatcactgt
601   taagtta
```

Figure 32

```
1     tgcgctatta ataattttcc ttatgataac tactatattt atagagacct atatgttttt
61    atttgtcatt tctttatact attctcttga ttttggggac gatagagatt gtcatgagaa
121   acagtacatt actaattaat aataaaggtg tgaacagaaa taagaagaaa tgaaaatact
181   aaaaaactat gcctacaatc tttcttatca attgttggtg atcatactcc ctatcattac
241   gactccctat gtaacgaggg tttttagttc tgacgattta ggaacgtatg gctactttag
301   ctccattgtt acctatttta ccttgcttgc aactcttggt gttgccaact acggtaccaa
361   agagatttca gcacatcgta aggaaattgg gaagaatttc tggggaattt attctctcca
421   gtttggtgca acttggctat ccattttgct ttatcttgcc ctttgtttct tatttacttc
481   aatgcaaaat ccggtagctt atatattggg attaagttta gtgtcaaaag gtttggatat
541   ttcttggtta tttcaaggtt tggaggattt tagaaagatt acagttcgga acatcactgt
601   taagtta
```

Figure 33

1	tgcgctatta	ataattttcc	ttatgataac	tactatattt	atagagacct	atatgttttt
61	atttgtcatt	tctttatact	attctcttga	ttttggggac	gatagagatt	gtcatgagaa
121	acagtacatt	actaattaat	aataaaggtg	tgaacagaaa	taagaagaaa	tgaaaatact
181	aaaaaattat	gcctacaatc	tttcttatca	attgttggtg	atcatactcc	ctatcattac
241	gactccctat	gtaacgaggg	tttttagttc	tgacgattta	ggaacgtatg	gctactttag
301	ctccattgtt	acctatttta	ccttgcttgc	aactcttggt	gttgccaact	acggtaccaa
361	agagatttca	gcacatcgta	aggaaattgg	gaagaatttc	tggggaattt	attctctcca
421	gtttggtgca	acttggctat	ccattttgct	ttatcttgcc	ctttgtttct	tatttacttc
481	aatgcaaaat	ccggtagctt	atatattggg	attaagttta	gtgtcaaaag	gtttggatat
541	ttcttggtta	tttcaaggtt	tggaggattt	tagaaagatt	acagttcgga	acatcactgt
601	taagtta					

Figure 34

```
1      tgcgctatta ataattttcc ttatgataac tactatattt atagagacct atatgttttt
61     atttgtcatt tctttatact attctcttga ttttggggac gatagagatt gtcatgagaa
121    acagtacatt actaattaat aataaaggtg tgaacagaaa taagaagaaa tgaaaatact
181    aaaaaattat gcctacaatc tttcttatca attgttggtg atcatactcc ctatcattac
241    gactccctat gtaacgaggg tttttagttc tgacgattta ggaacgtatg gctactttag
301    ctccattgtt acctatttta ccttgcttgc aactcttggt gttgccaact acggtaccaa
361    agagatttca gcacatcgta aggaaattgg gaagaatttc tggggaattt attctctcca
421    gtttggtgca acttggctat ccattttgct ttatcttgcc ctttgtttct tatttacttc
481    aatgcaaaat ccggtagctt atatattggg attaagttta gtgtcaaaag gtttggatat
541    ttcttggtta tttcaaggtt tggaggattt tagaaagatt acagttcgga acatcactgt
601    taagtta
```

Figure 35

```
1      tgcgctatta ataattttcc ttatgataac tactatattt atagagacct atatgttttt
61     atttgtcatt tctttatact attctcttga ttttggggac gatagagatt gtcatgagaa
121    acagtacatt actaattaat aataaaggtg tgaacagaaa taagaagaaa tgaaaatact
181    aaaaaactat gcctacaatc tttcttatca attgttggtg atcatactcc ctatcattac
241    gactccctat gtaacgaggg tttttagttc tgacgattta ggaacgtatg gctactttag
301    ctccattgtt acctatttta ccttgcttgc aactcttggt gttgccaact acggtaccaa
361    agagatttca gcacatcgta aggaaattgg gaagaatttc tggggaattt attctctcca
421    gtttggtgca acttggctat ccattttgct ttatcttgcc ctttgtttct tatttacttc
481    aatgcaaaat ccggtagctt atatattggg attaagttta gtgtcaaaag gtttggatat
541    ttcttggtta tttcaaggtt tggaggattt tagaaagatt acagttcgga acatcactgt
601    taagtta
```

Figure 36

```
1     tgcgctatta ataattttcc ttatgataac tactatattt atagagacct atatgttttt
61    atttgtcatt tctttatact attctcttga ttttggggac gatagagatt gtcatgagaa
121   acagtacatt actaattaat aataaaggtg tgaacagaaa taagaagaaa tgaaaatact
181   aaaaaactat gcctacaatc tttcttatca attgttggtg atcatactcc ctatcattac
241   gactccctat gtaacgaggg tttttagttc tgacgattta ggaacgtatg gctactttag
301   ctccattgtt acctatttta ccttgcttgc aactcttggt gttgccaact acggtaccaa
361   agagatttca gcacatcgta aggaaattgg gaagaatttc tggggaattt attctctcca
421   gtttggtgca acttggctat ccattttgct ttatcttgcc ctttgtttct tatttacttc
481   aatgcaaaat ccggtagctt atatattggg attaagttta gtgtcaaaag ggttggatat
541   ttcttggtta tttcaaggtt tggaggattt tagaaagatt acagttcgga acatcactgt
601   taagtta
```

Figure 37

```
1     tgatcatact ccctatcatt acgactccct atgtaacgag ggtcttttct tcggatgatt
61    tagggacgta tggttatttt aattccatcg ttacttattt tatcctctta gcgacgctag
121   gagttgctaa ctatgggacc aaggtcattt cagggcatcg aaagcaaatt caaaaaaact
181   ttttgggaat ctattctctg caattaggtg caacagttct ttctctgtcc ttgtatgctc
241   ttctttgtct aactcttccc tttatgcaaa atccggtagc ctatattcta ggcttgagtt
301   tagtttctaa aggtttagac atctcctggc tctttcaagg gttagaagat tttcgtaaaa
361   ttacggtcag aaatatcaca gtgaagctt
```

Figure 38

```
1     gagagtttgt acagtcactt actgaatcag tagaggggag aatcttgcct aatttaaaga
61    aaaacattgt ttacaatgtc ttatatcaga tcttagctgt aatagtaccg tttattacct
121   caccttactt agcgcgtgtg ttaggtgcag agcaaattgg agtttattct tttacttatt
181   ccattgcttt ttactttatg attctgtcca tgttgggaat ttctaattat gggaatcgga
241   caatggcaca ggtacgaaca agtagagaac atttgaatca agaattttcg aatatttacg
301   cggttcagtt gacgtgttca ctagtaatga ccgtctcata tttgatttat gcaacagtat
361   ttgtgaatag ttttcagatt gtagcctata tccaagtatt acatgtttta tcgtatgcaa
421   cagatgttag ttggtttttt tatggtcttg aagagtttcg tattacggtt gctaggaatt
481   catttgttaa gttattaact ttaatatcta tctttacatt tgtaaaaagc cctaatgata
541   tctatttata tacctttata atggcaggga gtaccctgct tggtcagttg attacatggc
601   cattttt
```

Figure 39

```
1     gagagtttgt acagttattt actgaatcag tagaggggag aatcttgcct agtttaaaga
61    aaaatattgt ttacaatgtc ttatatcaga tcttagctgt aatagtacca tttattacct
121   caccttactt agcgcgtgtg ttaggtgcag agcaaattgg agtttattct tttacttatt
181   ccattgcttt ttactttatg attctgtcca tgttggggat ttctaattat gggaatcgga
241   caatagcacg ggtacgaaca agtagagaac atttgaatca ggaattttcg aatatttacg
301   cggttcagtt gacgtgttca ctagtaatga ccatctcata tttgatttat gcaacagtat
361   ttgtgaatag ttttcagatt gtagcctata tccaagtatt acatgtttta tcgtatgcaa
421   cagatgttag ttggtttttt tatggtcttg aagagtttcg tattacggtt gctaggaatt
481   catttgttaa gttattaact ttaatatcta tctttacatt tgtaaaaagc cctaatgata
541   tctatttata tacctttata atggcagggg gtaccctgct tggtcagttg attacatggc
601   cattttt
```

Figure 40

```
1     gagagtttgt acagttattt actgaatcag tagaggggag aatcttgcct agtttaaaga
61    aaaatattgt ttacaatgtc ttatatcaga tcttagctgt aatagtacca tttattacct
121   caccttactt agcgcgtgtg ttaggtgcag agcaaattgg agtttattct tttacttatt
181   ccattgcttt ttactttatg attctgtcca tgttggggat ttctaattat gggaatcgga
241   caatagcacg ggtacgaaca agtagagaac atttgaatca ggaattttcg aatatttacg
301   cggttcagtt gacgtgttca ctagtaatga ccatctcata tttgatttat gcaacagtat
361   ttgtgaatag ttttcagatt gtagcctata tccaagtatt acatgtttta tcgtatgcaa
421   cagatgttag ttggtttttt tatggtcttg aagagtttcg tattacggtt gctaggaatt
481   catttgttaa gttattaact ttaatatcta tctttacatt tgtaaaaagc cctaatgata
541   tctatttata tacctttata atggcagggg gtaccctgct tggtcagttg attacatggc
601   catttttt
```

Figure 41

```
1     gagagtttgt acagtcactt actgaatcag tagaggggag aatcttgcct aatttaaaga
61    aaaacattgt ttacaatgtc ttatatcaga tcttagctgt aatagtaccg tttattacct
121   caccttactt agcgcgtgtg ttaggtgcag agcaaattgg agtttattct tttacttatt
181   ccattgcttt ttactttatg attctgtcca tgttgggaat ttctaattat gggaatcgga
241   caatagcaca ggtacgaaca agtagagaac atttgaatca agaattttcg aatatttacg
301   cagttcagtt gacgtgttca ctagtaatga ccgtctcata tttgatttat gcaacagtat
361   ttgtgaatag ttttcagatt gtagcctata tccaagtatt acatgtttta tcgtatgcaa
421   cagatgttag ttggtttttt tatggtcttg aagagtttcg tattacggtt gctaggaatt
481   catttgttaa gttattaact ttaatatcta tctttacatt tgtaaaaagc cctaatgata
541   tctatttata tacctttata atggcaggga gtaccctgct tggtcagttg attacatggc
601   aattttt
```

Figure 42

1	gagagtttgt	acagtcactt	actgaatcag	tagaggggag	aatcttgcct	aatttaaaga
61	aaaacattgt	ttacaatgtc	ttatatcaga	tcttagctgt	aatagtaccg	tttattacct
121	caccttactt	agcgcgtgtg	ttaggtgcag	agcaaattgg	agtttattct	tttacttatt
181	ccattgcttt	ttactttatg	attctgtcca	tgttgggaat	ttctaattat	gggaatcgga
241	caatagcaca	ggtacgaaca	agtagagaac	atttgaatca	agaattttcg	aatatttacg
301	cagttcagtt	gacgtgttca	ctagtaatga	ccgtctcata	tttgatttat	gcaacagtat
361	ttgtgaatag	ttttcagatt	gtagcctata	tccaagtatt	acatgtttta	tcgtatgcaa
421	cagatgttag	ttggtttttt	tatggtcttg	aagagtttcg	tattacggtt	gctaggaatt
481	catttgttaa	gttattaact	ttaatatcta	tctttacatt	tgtaaaaagc	cctaatgata
541	tctatttata	tacctttata	atggcaggga	gtaccctgct	tggtcagttg	attacatggc
601	aattttt					

Figure 43

```
1     gagagtttgt acagtcactt actgaatcag tagaggggag aatcttgcct aatttaaaga
61    aaaacattgt ttacaatgtc ttatatcaga tcttagctgt aatagtaccg tttattacct
121   caccttactt agcgcgtgtg ttaggtgcag agcaaattgg agtttattct tttacttatt
181   ccattgcttt ttactttatg attctgtcca tgttgggaat ttctaattat gggaatcgga
241   caatagcaca ggtacgaaca agtagagaac atttgaatca agaattttcg aatatttacg
301   cagttcagtt gacgtgttca ctagtaatga ccgtctcata tttgatttat gcaacagtat
361   ttgtgaatag ttttcagatt gtagcctata tccaagtatt acatgtttta tcgtatgcaa
421   cagatgttag ttggtttttt tatggtcttg aagagtttcg tattacggtt gctaggaatt
481   catttgttaa gttattaact ttaatatcta tctttacatt tgtaaaaagc cctaatgata
541   tctatttata tacctttata atggcaggga gtaccctgct tggtcagttg attacatggc
601   aattttt
```

Figure 44

```
1       gagagtttgt acagtcactt actgaatcag tagaggggag aatcttgcct aatttaaaga
61      aaaacattgt ttacaatgtc ttatatcaga tcttagctgt aatagtaccg tttattacct
121     caccttactt agcgcgtgtg ttaggtgcag agcaaattgg agtttattct tttacttatt
181     ccattgcttt ttactttatg attctgtcca tgttgggaat ttctaattat gggaatcgga
241     caatagcaca ggtacgaaca agtagagaac atttgaatca agaattttcg aatatttacg
301     cagttcagtt gacgtgttca ctagtaatga ccgtctcata tttgatttat gcaacagtat
361     ttgtgaatag ttttcagatt gtagcctata tccaagtatt acatgtttta tcgtatgcaa
421     cagatgttag ttggtttttt tatggtcttg aagagtttcg tattacggtt gctaggaatt
481     catttgttaa gttattaact ttaatatcta tctttacatt tgtaaaaagc cctaatgata
541     tctatttata tacctttata atggcaggga gtaccctgct tggtcagttg attacatggc
601     aattttt
```

Figure 45

```
1     gagagtttgt acagtcactt actgaatcag tagaggggag aatcttgcct aatttaaaga
61    aaaacattgt ttacaatgtc ttatatcaga tcttagctgt aatagtaccg tttattacct
121   caccttactt agcgcgtgtg ttaggtgcag agcaaattgg agtttattct tttacttatt
181   ccattgcttt ttactttatg attctgtcca tgttgggaat ttctaattat gggaatcgga
241   caatagcaca ggtacgaaca agtagagaac atttgaatca agaattttcg aatatttacg
301   cagttcagtt gacgtgttca ctagtaatga ccgtctcata tttgatttat gcaacagtat
361   ttgtgaatag ttttcagatt gtagcctata tccaagtatt acatgtttta tcgtatgcaa
421   cagatgttag ttggtttttt tatggtcttg aagagtttcg tattacggtt gctaggaatt
481   catttgttaa gttattaact ttaatatcta tctttacatt tgtaaaaagc cctaatgata
541   tctatttata tacctttata atggcaggga gtaccctgct tggtcagttg attacatggc
601   aattttt
```

Figure 46

```
1      gagagtttgt acagtcactt actgaatcag tagaggggag aatcttgcct aatttaaaga
61     aaaacattgt ttacaatgtc ttatatcaga tcttagctgt aatagtaccg tttattacct
121    caccttactt agcgcgtgtg ttaggtgcag agcaaattgg agtttattct tttacttatt
181    ccattgcttt ttactttatg attctgtcca tgttgggaat ttctaattat gggaatcgga
241    caatagcaca ggtacgaaca agtagagaac atttgaatca agaattttcg aatatttacg
301    cagttcagtt gacgtgttca ctagtaatga ccgtctcata tttgatttat gcaacagtat
361    ttgtgaatag ttttcagatt gtagcctata tccaagtatt acatgtttta tcgtatgcaa
421    cagatgttag ttggtttttt tatggtcttg aagagtttcg tattacggtt gctaggaatt
481    catttgttaa gttattaact ttaatatcta tctttacatt tgtaaaaagc cctaatgata
541    tctatttata tacctttata atggcaggga gtaccctgct tggtcagttg attacatggc
601    aattttt
```

Figure 47

```
1      caggatatag atgcatgtta gattagatgg tttgctggac tatatatttc tatttagtgt
61     gattattact tgtaatacta tgtattcaac tagtcaagga tttgatggac tagggaaatg
121    ggcgactctg ttacttgtgg tatcagtttt tctgaaattg cttatctcta gaatatctat
181    gaaggcaatc aatgtgattg tgtcgcgttc tttaatattt atattaatta ttctactcat
241    agtaatatta aatggtttta agatttctga gacaagtttc gtctattatt ttgtattatt
301    tccgattttt atgatgattt tgcagatgta ctatgatgtt aatgaaatcg caaatctgat
361    acggaaattt gttcgtataa tatttctttt agcaattggc tctctcctat tttggcttat
421    tggtagtgta tttcatatta tatccccaac ggtttatgtg ttgaattatt ggaatggtgg
481    gggaatagta gaagggtact ataatcttca ttttgaagca caaaaaatag agattttggg
541    ggcgata
```

Figure 48

```
1      caggatatag atgcatgtta gattagatgg tttgctggac tatatatttc tatttagtgt
61     gattattact tgtaatacta tgtattcaac tagtcaagga tttgatggac tagggaaatg
121    ggcgactctg ttacttgtgg tatcagttat tctgaaattg cttatctcta gaatatctat
181    gaaggcaatc aatgtgattg tgtcgcgttc tttaatattt atattaatta ttctactcat
241    agtaatatta aatggtttta agatttctga gacaagtttc gtctattatt ttgtattatt
301    tccgattttt atgatgattt tgcagatgta ctatgatgtt aatgaaatcg caaatctgat
361    acggaaattt gttcgtataa tatttctttt agcaattggc tctctcctat tttggcttat
421    tggtagtgta tttcatatta tatccccaac ggtttatgtg ttgaattatt ggaatggtgg
481    gggaatagta gaagggtact ataatcttca ttttgaagca caaaaaatag agattttggg
541    ggcgata
```

Figure 49

```
1     caggatatag atgcatgtta gattagatgg tttgctggac tatatatttc tatttagtgt
61    gattattact tgtaatacta tgtattcaac tagtcaagga tttgatggac tagggaaatg
121   ggcgactctg ttacttgtgg tatcagttat tctgaaattg cttatctcta gaatatctat
181   gaaggcaatc aatgtgattg tgtcgcgttc tttaatattt atattaatta ttctactcat
241   agtaatatta aatggtttta agatttctga gacaagtttc gtctattatt ttgtattatt
301   tccgattttt atgatgattt tgcagatgta ctatgatgtt aatgaaatcg caaatctgat
361   acggaaattt gttcgtataa tatttctttt agcaattggc tctctcctat tttggcttat
421   tggtagtgta tttcatatta tatccccaac ggtttatgtg ttgaattatt ggaatggtgg
481   gggaatagta gaagggtact ataatcttca ttttgaagca caaaaaatag agattttggg
541   ggcgata
```

Figure 50

1	caggatatag	atgcatgtta	gattagatgg	tttgctggac	tatatatttc	tatttagtgt
61	gattattact	tgtaatacta	tgtattcaac	tagtcaagga	tttgatggac	tagggaaatg
121	ggcgactctg	ttacttgtgg	tatcagttat	tctgaaattg	cttatctcta	gaatatctat
181	gaaggcaatc	aatgtgattg	tgtcgcgttc	tttaatattt	atattaatta	ttctactcat
241	agtaatatta	aatggtttta	agatttctga	gacaagtttc	gtctattatt	ttgtattatt
301	tccgattttt	atgatgattt	tgcagatgta	ctatgatgtt	aatgaaatcg	caaatctgat
361	acggaaattt	gttcgtataa	tatttctttt	agcaattggc	tctctcctat	tttggcttat
421	tggtagtgta	tttcatatta	tatccccaac	ggtttatgtg	ttgaattatt	ggaatggtgg
481	gggaatagta	gaagggtact	ataatcttca	ttttgaagca	caaaaaatag	agattttggg
541	ggcgata					

Figure 51

```
1     caggatatag atgcatgtta gattagatgg tttgctggac tatatatttc tatttagtgt
61    gattattact tgtaatacta tgtattcaac tagtcaagga tttgatggac tagggaaatg
121   ggcgactctg ttacttgtgg tatcagtttt tctgaaattg cttatctcta gaatatctat
181   gaaggcaatc aatgtgattg tgtcgcgttc tttaatattt atattaatta ttctactcat
241   agtaatatta aatggtttta agatttctga gacaagtttc gtctattatt ttgtattatt
301   tccgattttt atgatgattt tgcagatgta ctatgatgtt aatgaaatcg caaatctgat
361   acggaaattt gttcgtataa tatttctttt agcaattggc tctctcctat tttggcttat
421   tggtagtgta tttcatatta tatccccaac ggtttatgtg ttgaattatt ggaatggtgg
481   gggaatagta gaagggtact ataatcttca ttttgaagca caaaaaatag agattttggg
541   ggcgata
```

Figure 52

1	caggatatag	atgcatgtta	gattagatgg	tttgctggac	tatatatttc	tatttagtgt
61	gattattact	tgtaatacta	tgtattcaac	tagtcaagga	tttgatggac	tagggaaatg
121	ggcgactctg	ttacttgtgg	tatcagttat	tctgaaattg	cttatctcta	gaatatctat
181	gaaggcaatc	aatgtgattg	tgtcgcgttc	tttaatattt	atattaatta	ttctactcat
241	agtaatatta	aatggtttta	agatttctga	gacaagtttc	gtctattatt	ttgtattatt
301	tccgattttt	atgatgattt	tgcagatgta	ctatgatgtt	aatgaaatcg	caaatctgat
361	acggaaattt	gttcgtataa	tatttctttt	agcaattggc	tctctcctat	tttggcttat
421	tggtagtgta	tttcatatta	tatccccaac	ggtttatgtg	ttgaattatt	ggaatggtgg
481	gggaatagta	gaagggtact	ataatcttca	ttttgaagca	caaaaaatag	agattttggg
541	ggcgata					

Figure 53

```
1     caggatatag atgcatgtta gattagatgg tttgctggac tatatatttc tatttagtgt
61    gattattact tgtaatacta tgtattcaac tagtcaagga tttgatggac tagggaaatg
121   ggcgactctg ttacttgtgg tatcagtttt tctgaaattg cttatctcta gaatatctat
181   gaaggcaatc aatgtgattg tgtcgcgttc tttaatattt atattaatta ttctactcat
241   agtaatatta aatggtttta agatttctga gacaagtttc gtctattatt ttgtattatt
301   tccgattttt atgatgattt tgcagatgta ctatgatgtt aatgaaatcg caaatctgat
361   acggaaattt gttcgtataa tatttctttt agcaattggc tctctcctat tttggcttat
421   tggtagtgta tttcatatta tatccccaac ggtttatgtg ttgaattatt ggaatggtgg
481   gggaatagta gaagggtact ataatcttca ttttgaagca caaaaaatag agattttggg
541   ggcgata
```

Figure 54

1	caggatatag	atgcatgtta	gattagatgg	tttgctggac	tatatatttc	tatttagtgt
61	gattattact	tgtaatacta	tgtattcaac	tagtcaagga	tttgatggac	tagggaaatg
121	ggcgactctg	ttacttgtgg	tatcagtttt	tctgaaattg	cttatctcta	gaatatctat
181	gaaggcaatc	aatgtgattg	tgtcgcgttc	tttaatattt	atattaatta	ttctactcat
241	agtaatatta	aatggtttta	agatttctga	gacaagtttc	gtctattatt	ttgtattatt
301	tccgattttt	atgatgattt	tgcagatgta	ctatgatgtt	aatgaaatcg	caaatctgat
361	acggaaattt	gttcgtataa	tatttctttt	agcaattggc	tctctcctat	tttggcttat
421	tggtagtgta	tttcatatta	tatccccaac	ggtttatgtg	ttgaattatt	ggaatggtgg
481	gggaatagta	gaagggtact	ataatcttca	ttttgaagca	caaaaaatag	agattttggg
541	ggcgata					

Figure 55

1	tttttagaac	gtactcattt	atttaaaagg	aagtaatagt	gaaatttaaa	tttaaattta
61	atccaatcgc	gatactgtat	atattgctag	tatacttaga	gttagctaca	gataggcaac
121	atctgtatcc	tgtaacgtac	atgacaaaat	attatattgg	tattttaatc	actgtgttgt
181	ttgttttgtt	attagtaggc	cgtgggaagc	ttatttttgt	taataaaaaa	ttattatatc
241	ttgctaagat	attagctata	ccaacaattg	ttcttttcct	gtactcagtc	ttactagacg
301	taatgaaccc	agttgaattt	gatggatatt	ttagtaggtt	atcaagtacg	actatttttg
361	gtttgttagc	tatctttcaa	gctatagttg	tttttcaatt	ttttggacaa	aaagtagtag
421	attacacttt	tacagctatc	tccctcagct	acttaaccag	tatcattgtt	gcctttaggc
481	agggaggact	tagtcaattt	atcttgatgc	taacagatga	tagtttcaat	ggttcggtac
541	tagaaat					

Figure 56

1	tttttagaac	gtactcattt	atttaaaagg	aagtaatagt	gaaatttaaa	tttaaattta
61	atccaatcgc	gatactgtat	atattgctag	tatacttaga	gttagctaca	gataggcaac
121	atctgtatcc	tgtaacgtac	atgacaaaat	attatattgg	tattttaatc	actgtgttgt
181	ttgttttgtt	attagtaggc	cgtgggaagc	ttatttttgt	taataaaaaa	ttattatatc
241	ttgctaagat	attagctata	ccaacaattg	ttcttttcct	gtactcagtc	ttactagacg
301	taatgaaccc	agttgaattt	gatggatatt	ttagtaggtt	atcaagtacg	actatttttg
361	gtttgttagc	tatctttcaa	gctatagttg	tttttcaatt	ttttggacaa	aaagtagtag
421	attacacttt	tacagctatc	tccctcagct	acttaaccag	tatcattgtt	gcctttaggc
481	agggaggact	tagtcaattt	atcttgatgc	taacagatga	tagtttcaat	ggttcggtac
541	tagaaat					

Figure 57

```
1     tttttagaac atactcattt atttaaaagg aaataatagt gaaatttaaa tttaatccaa
61    tcgcgatact gtatatattg ctagtatact tagagttggc tacagatagg caacatctgt
121   atcctgtaac gtacatgaca aaatattata ttggtatttt aatcattgtg ttgtttgttt
181   tattattagt aggccgtggg aagcttattt ttgttaataa aaaattatta tatcttgcta
241   agatattagc tataccaaca attgttcttt tcctgtactc agtcttacta gacgtaatga
301   acccagttga atttaatgga tattttagta gattatcaag tacgactatt tttggtttgt
361   tagctatctt tcaagctata gttgtttttc aatttttggg acaaaaagta gtagattaca
421   cttttacagc tatctccctc agctacttaa ccagtatcat tgttgccttt aggcagggag
481   gacttagtca atttatcttg atactaacag atgatagttt caatggttcg gtactagaaa
541   t
```

Figure 58

```
1     tttttagaac atactcattt atttaaaagg aaataatagt gaaatttaaa tttaatccaa
61    tcgcgatact gtatatattg ctagtatact tagagttggc tacagatagg caacatctgt
121   atcctgtaac gtacatgaca aaatattata ttggtatttt aatcattgtg ttgtttgttt
181   tattattagt aggccgtggg aagcttattt ttgttaataa aaaattatta tatcttgcta
241   agatattagc tataccaaca attgttcttt tcctgtactc agtcttacta gacgtaatga
301   acccagttga atttaatgga tattttagta gattatcaag tacgactatt tttggtttgt
361   tagctatctt tcaagctata gttgtttttc aatttttttgg acaaaaagta gtagattaca
421   cttttacagc tatctccctc agctacttaa ccagtatcat tgttgccttt aggcagggag
481   gacttagtca atttatcttg atactaacag atgatagttt caatggttcg gtactagaaa
541   t
```

Figure 59

```
1     tttttagaac atactcattt atttaaaagg aaataatagt gaaatttaaa tttaatccaa
61    tcgcgatact gtatatattg ctagtatact tagagttggc tacagatagg caacatctgt
121   atcctgtaac gtacatgaca aaatattata ttggtatttt aatcattgtg ttgtttgttt
181   tattattagt aggccgtggg aagcttattt ttgttaataa aaaattatta tatcttgcta
241   agatattagc tataccaaca attgttcttt tcctgtactc agtcttacta gacgtaatga
301   acccagttga atttaatgga tattttagta gattatcaag tacgactatt tttggtttgt
361   tagctatctt tcaagctata gttgtttttc aatttttttgg acaaaaagta gtagattaca
421   cttttacagc tatctccctc agctacttaa ccagtatcat tgttgccttt aggcagggag
481   gacttagtca atttatcttg atactaacag atgatagttt caatggttcg gtactagaaa
541   t
```

Figure 60

1	tttttagaac	atactcattt	atttaaaagg	aaataatagt	gaaatttaaa	tttaatccaa
61	tcgcgatact	gtatatattg	ctagtatact	tagagttggc	tacagatagg	caacatctgt
121	atcctgtaac	gtacatgaca	aaatattata	ttggtatttt	aatcattgtg	ttgtttgttt
181	tattattagt	aggccgtggg	aagcttattt	ttgttaataa	aaaattatta	tatcttgcta
241	agatattagc	tataccaaca	attgttcttt	tcctgtactc	agtcttacta	gacgtaatga
301	acccagttga	atttaatgga	tattttagta	gattatcaag	tacgactatt	tttggtttgt
361	tagctatctt	tcaagctata	gttgtttttc	aatttttttgg	acaaaaagta	gtagattaca
421	cttttacagc	tatctccctc	agctacttaa	ccagtatcat	tgttgccttt	aggcagggag
481	gacttagtca	atttatcttg	atactaacag	atgatagttt	caatggttcg	gtactagaaa
541	t					

Figure 61

```
1     tttttagaac atactcattt atttaaaagg aagtaatagt gaaatttaaa tttaatccaa
61    tcgcgatact gtatatattg ctagtatact tagagttggc tacagatagg caacatctgt
121   atcctgtaac gtacatgaca aaatattata ttggtatttt aatcattgtg ttgtttgttt
181   tattattagt aggccgtggg aagcttattt ttgttaataa aaaattatta tatcttgcta
241   agatattagc tataccaaca attgttcttt tcctgtactc agtcttacta gacgtaatga
301   acccagttga atttaatgga tattttagta ggttatcaag tacgactatt tttggtttgt
361   tagctatctt tcaagctata gttgtttttc aatttttttgg acaaaaagta gtagattaca
421   cttttacagc tatctccctc agctacttaa ccagtatcat tgttgccttt aggcagggag
481   gacttagtca atttatcttg atactaacag atgatagttt caatggttcg gtactagaaa
541   t
```

Figure 62

1	tttttagaac	atactcattt	atttaaaagg	aaataatagt	gaaatttaaa	tttaatccaa
61	tcgcgatact	gtatatattg	ctagtatact	tagagttggc	tacagatagg	caacatctgt
121	atcctgtaac	gtacatgaca	aaatattata	ttggtatttt	aatcattgtg	ttgtttgttt
181	tattattagt	aggccgtggg	aagcttattt	ttgttaataa	aaaattatta	tatcttgcta
241	agatattagc	tataccaaca	attgttcttt	tcctgtactc	agtcttacta	gacgtaatga
301	acccagttga	atttaatgga	tattttagta	gattatcaag	tacgactatt	tttggtttgt
361	tagctatctt	tcaagctata	gttgtttttc	aatttttttgg	acaaaaagta	gtagattaca
421	cttttacagc	tatctccctc	agctacttaa	ccagtatcat	tgttgccttt	aggcagggag
481	gacttagtca	atttatcttg	atactaacag	atgatagttt	caatggttcg	gtactagaaa
541	t					

Figure 63

```
1      tttttagaac atactcattt atttaaaagg aaataatagt gaaatttaaa tttaatccaa
61     tcgcgatact gtatatattg ctagtatact tagagttggc tacagatagg caacatctgt
121    atcctgtaac gtacatgaca aaatattata ttggtatttt aatcattgtg ttgtttgttt
181    tattattagt aggccgtggg aagcttattt ttgttaataa aaaattatta tatcttgcta
241    agatattagc tataccaaca attgttcttt tcctgtactc agtcttacta gacgtaatga
301    acccagttga atttaatgga tattttagta gattatcaag tacgactatt tttggtttgt
361    tagctatctt tcaagctata gttgtttttc aatttttttgg acaaaaagta gtagattaca
421    cttttacagc tatctccctc agctacttaa ccagtatcat tgttgccttt aggcagggag
481    gacttagtca atttatcttg atactaacag atgatagttt caatggttcg gtactagaaa
541    t
```

Figure 64

```
                  + Mct 4
               +-68
               !  + Mcst 10A-q
               !
               !        + Mcst 33F-g
            +-67     +-65
            !  !     !  + Mct 33A
            !  !  +-64
            !  !  !  !  + Mct 35B
            !  +-66  +-63
         +-62     !     + Mcst 17F-35B
         !  !     !
         !  !     + Mcst 17F-c
         !  !
      +-69  !  + Mct 34
      !  !  +-61
      !  !     + Mct 17A
   +-60  !
   !  !  + Mct 35F
   !  !
   !  !  + Mct 37
+-56  +-59
!  !     + Mcst 23A-ca
!  !
!  !  + Mcst 6B-g
!  +-57
!     + Mcst 6A-6B-g
!
!
!                                         +.Mct 7C
!                                      +-49
!                                   +-50  + Mcst 15A-cal
!                                   !  !
!                                +-51  + Mct 9V
!                                !  !
!                                !  + Mcst 6B-c
!                             +-48
!                             !  !     + Mct 22F
!                             !  !  +-46
!                          +-52  +-47  + Mcst 15B-22F
!                          !  !     !
!                          !  !     + Mct 22A
!                          !  !
!                       +-45  + Mct 21
!                       !  !
!                       !  !     + Mcst 33F-q
!                       !  !  +-43
!                    +-53  +-44  + Mct 33B
!                    !  !     !
!                    !  !     + Mcst 11A-nz
!                 +-54  !
!                 !  !  + Mct 10F
!              +-55  !
!              !  !  + Mct 12F
!              !  !
!              !  + Mcst 2-q
!              !
!           +-42           + Mct 42
!           !  !        +-39
!           !  !     +-40  + Mct 31
!           !  !     !  !
!           !  !  +-41  + Mcst 5-c
!           !  !  !  !
```

Continue next page

```
            !           !  +-38  + Mct 8
            !           !     !
            !           !     !  + Mct 41F
            !           !     +-37
         +-58           !        + Mcst 2-g
         !  !           !
         !  !           !                          + Mcst 23F-g
         !  !           !                       +-30
         !  !           !                    +-31  + Mcst 10A-23F
         !  !           !                    !  !
         !  !           !                    !  + Mct 14-g
         !  !        +-36                 +-29
         !  !        !  !                 !  !     + Mcst 15C-q
         !  !        !  !                 !  !  +-27
         !  !        !  !              +-32  +-28  + Mcst 15B-q
         !  !        !  !              !  !     !
         !  !        !  !              !  !     + Mcst 15C-ca
         !  !        !  !           +-26  !
         !  !        !  !           !  !  + Mct 29
         !  !        !  !           !  !
         !  !        !  !        +-33  !  + Mcst 23F-23A
         !  !        !  !        !  !  +-25
         !  !        !  !        !  !  + Mcst 23A-23F
         !  !     +-22  !     +-34  !
         !  !     !  !  !     !  !  + Mct 7F
         !  !     !  !  !  +-35  !
         !  !     !  !  !  !  !  + Mcst 14-c
         !  !     !  !  !  !  !
         !  !     !  !  +-24  + Mcst 5-q
         !  !     !  !     !
         !  !     !  !     !  + Mct 16F
         !  !     !  !     +-23
         !  !     !  !        + Mcst 15B-c
         !  !     !  !
      +-70  !     !  !     + Mct 20
      !  !  !     !  !  +-20
      !  !  !     !  +-21  + Mct 13
      !  !  !  +-19     !
      !  !  !  !  !     + Mct 9N
      !  !  !  !  !
      !  !  !  !  !                 + Mct 18C
      !  !  !  !  !              +-14
      !  !  !  !  !           +-15  + Mct 18B
      !  !  !  !  !           !  !
      !  !  !  !  !        +-16  + Mct 19F
      !  !  !  !  !        !  !
      !  !  !  !  !     +-17  + Mct 18F
      !  !  !  !  !     !  !
      !  !  !  !  !  +-18  + Mct 1
      !  !  +--8  !  !  !
      !  !     !  !  !  + Mct 18A
      !  !     !  +-13
+-----7  !     !     !     + Mcst 23F-c
!     !  !     !     !  +-12
!     !  !     !     !  !  + Mcst 15A-ca2
!     !  !     !     +-11
!     !  !     !        !     + Mcst 6A-ca
!     !  !     !        !  +--9
!     !  !     !        +-10  + Mcst 6A-c
!     !  !     !           !
!     !  !     !           + Mcst 6A-g
!     !  !     !
!     !  !     +-- Mcst 11A-q
```

Continue next page

```
           !    !  !
        +--2    !  +--- Mct 3
        !  !    !
        !  !    !        + Mcst 6B-q
        !  !    !     +--5
        !  !    !  +--6  + Mcst 6A-6B-q
        !  !    !  !  !
        !  !    +--4  + Mcst 11B
        !  !       !
  +----71  !       !  + Mct 23B
  !     !  !       +--3
  !     !  !          + Mct 19A
  !     !  !
  !     !  !         + Mct 38
-72     !  +----------1
  !     !            + Mct 25
  !     !
  !     +----------Nontypable-nz
  !
  +----------------Nontypable-ca
```

Figure 65

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/AU04/000480

International filing date: 13 April 2004 (13.04.2004)

Document type: Certified copy of priority document

Document details: Country/Office: AU
Number: 2003901717
Filing date: 10 April 2003 (10.04.2003)

Date of receipt at the International Bureau: 15 March 2005 (15.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse